

How Punishment Severity Affects Jury Verdicts

Author(s): Anna Bindler and Randi Hjalmarsson

Source: *American Economic Journal: Economic Policy*, November 2018, Vol. 10, No. 4 (November 2018), pp. 36-78

Published by: American Economic Association

Stable URL: <https://www.jstor.org/stable/10.2307/26529053>

**REFERENCES**

Linked references are available on JSTOR for this article:

[https://www.jstor.org/stable/10.2307/26529053?seq=1&cid=pdf-reference#references\\_tab\\_contents](https://www.jstor.org/stable/10.2307/26529053?seq=1&cid=pdf-reference#references_tab_contents)

You may need to log in to JSTOR to access the linked references.

---

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



JSTOR

American Economic Association is collaborating with JSTOR to digitize, preserve and extend access to *American Economic Journal: Economic Policy*

## How Punishment Severity Affects Jury Verdicts: Evidence from Two Natural Experiments<sup>†</sup>

By ANNA BINDLER AND RANDI HJALMARSSON\*

*This paper studies the effect of punishment severity on jury decision making using archival data from London's Old Bailey Criminal Court from 1772 to 1871. We exploit two natural experiments in English history, resulting in sharp decreases in punishment severity: the offense-specific abolition of capital punishment and the temporary halt of penal transportation during the American Revolution. Using difference-in-differences to study the former and a pre-post design for the latter, we find a large, significant, and permanent impact on jury behavior: juries are more likely to convict overall and across crime categories. Moreover, the effect size differs with defendants' gender. (JEL K41, K42, N43)*

More than 50 years ago, President Lyndon B. Johnson declared a war on crime. Numerous policies have since been implemented that have increased expected punishment and are responsible for the dramatic—almost four-fold—growth in the US state and federal prison populations.<sup>1</sup> A new era of reforms aiming to reverse this “tough on crime” attitude by decreasing sentence severity is currently being ushered in, resulting in a (slowly) falling prison population. These reforms range from abolishing or reducing mandatory minimums to abolishing the death penalty, which still exists in 31 states today.<sup>2</sup> The goals of the sanction setting policymaker (in both sanction regimes) are multifaceted, including crime control, proportionate punishment, and a fair application of justice. However, the policymaker's ability to achieve these goals depends on the willingness of many other agents—police, prosecutors, juries, and judges—throughout the criminal justice system to implement

\*Bindler: Department of Economics, University of Gothenburg, Box 640, 40530 Gothenburg, Sweden (email: [anna.bindler@economics.gu.se](mailto:anna.bindler@economics.gu.se)); Hjalmarsson: Department of Economics, University of Gothenburg, Box 640, 40530 Gothenburg, Sweden (email: [randi.hjalmarsson@economics.gu.se](mailto:randi.hjalmarsson@economics.gu.se)). This paper would not have been possible without the tremendous efforts of our research assistants Michael Bekele and Srinidhi Srinivasan, the generous help with the data extraction by Florin Maican, and the financial support for Hjalmarsson of the Foundation for Economic Research in West Sweden (2250-242 334 602), and Vetenskapsrådet, the Swedish Research Council, Grants for Distinguished Young Researchers (446-2014-1735). We thank Daniel Klerman, Mikael Lindahl, Ines Helm, Tim Hitchcock, and seminar participants at the Conference for Empirical Legal Studies (2016), Bergen Stavanger Workshop (2017), WinE Retreat and EEA Annual Conference (2017), RES Annual Conference (2017), Youth Crime and Public Policy Interventions Conference Surrey (2017), University of Gothenburg, University of Oslo, Tillburg, Rotterdam, Stockholm University, and CERNA MINES Paris Tech for helpful comments.

<sup>†</sup>Go to <https://doi.org/10.1257/pol.20170214> to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

<sup>1</sup> See <https://www.bjs.gov/index.cfm?ty=kfdetail&iid=487> (accessed June 6, 2016).

<sup>2</sup> See <http://www.deathpenaltyinfo.org/states-and-without-death-penalty>.

these sanctions. This dichotomy between the policymaker and policy executors may be especially apparent in circumstances where the latter perceive a sanction to be particularly disproportionate.

Ultimately, one wants to know how these policies affect crime. In the economic model of crime (Becker 1968), an individual's decision to engage in crime is a function of expected punishment, which is in turn a function of the chances of getting caught, charged, and convicted, as well as the severity of punishment. Thus, the policymaker only directly controls the last of these inputs into expected punishment, while the reaction of the police, prosecutor, and judge or jury, respectively, to the policy change may also indirectly impact the behavior of a potential criminal through their effect on expected punishment.

This paper studies the reaction of one such agent—the jury—to criminal justice policies. Specifically, we capitalize on two natural experiments in English history associated with large and sharp changes in punishment severity—the offense-specific abolition of capital punishment in the 1800s and the temporary and unexpected halt of transportation during the American Revolution. Do changes in punishment severity affect jury decision making and the chance of conviction? Moreover, are juries impacted by punishment severity in a way that is unequally applied across defendants?

There is a growing empirical literature that studies how expected punishment may impact the behavior of other criminal justice agents, including parole boards, judges, and especially prosecutors. For instance, Bjerk (2005) finds that prosecutors react to three-strikes laws by lowering the charge to a misdemeanor when conviction of the original felony would lead to a third strike.<sup>3</sup> Yet, to the best of our knowledge, no study has looked at this final, high-stakes stage of the judicial process—the conviction.

A jury's job is to determine whether the facts of the case prove beyond reasonable doubt the defendant's guilt; the jury's evaluation should not be affected by factors external to the evidence of the case. Whether this holds in practice has been recently studied empirically with respect to jury demographics or contemporaneous media coverage during a trial.<sup>4</sup> There is limited research, however, on the role of potential punishment, which also should not factor into the jury's decision, and the existing research is generally unable to disentangle the effect of the severity of the punishment from the severity of the offense; Devine's (2012) review of the so-called "severity-leniency hypothesis" finds that no firm conclusions can be made.<sup>5</sup>

A related concept is that of "jury nullification": does the jury take the law into their own hands based on their own ethical beliefs, for instance by acquitting a defendant for whom the facts prove guilt beyond reasonable doubt? While we are

<sup>3</sup> Tonry (1992) reviews the impact of mandatory sentencing on lawyer and judge actions to avoid (nullify) the impact of these laws. Other studies of the effect of sentences on discretionary behavior include Bushway, Owens, and Piehl (2012) for judges; Ulmer, Kurlychek, and Kramer (2007) and Starr and Rehavi (2013) for prosecutors; and Lacasse and Payne (1999) for plea bargaining.

<sup>4</sup> See, for instance, Anwar, Bayer, and Hjalmarsson (2012, 2016), Lehmann and Blair Smith (2013), and Lee (2017) for studies of juror demographics and Philippe and Ouss (forthcoming) for the role of the media.

<sup>5</sup> Experimental studies of the severity-leniency hypothesis include Vidmar (1972) and Freedman et al. (1994). Nonexperimental studies (e.g., Devine et al. 2004 and Flower 2008) also proxy for expected punishment with charge severity and use small samples.

unaware of empirical studies of this question, there is anecdotal evidence of jury nullification throughout history, including defendants charged with helping slaves escape or Vietnam War protesters.<sup>6</sup> There is also clear anecdotal evidence of jury nullification with respect to capital punishment in nineteenth century England (Tonry 1992). Schefflin and Van Dyke (1980, 71) summarize this evidence: “At the beginning of the nineteenth century in England, persons convicted of violating any one of over 230 crimes were automatically sentenced to death. Jurors refused to convict because, although they recognized the defendant’s conduct as criminal, they could not condemn someone to death for an offense that they felt did not warrant such a penalty.” Yet, there is no empirical evidence regarding the degree to which this occurred and, perhaps more importantly, whether it was applied differentially across case and defendant characteristics.

To the best of our knowledge, this is the first study of the causal effect of changes in punishment severity on jury decision making using a quasi-experimental research design. Our identification strategy is unique to this literature in that we capitalize on changes in sentencing laws that decrease punishment severity for a given criminal offense. This contrasts previous research that asks whether juries are less likely to convict defendants charged with more serious offenses (e.g., robbery versus burglary), where one cannot disentangle the differential punishment severity across offenses from the differential case characteristics and/or evidence. In addition, our focus on a sharp reduction in punishment contrasts most (if not all) previous literature focused on how increases in sentence severity affect discretionary behavior.

England in the eighteenth and nineteenth centuries provides a colorful context during which to study changing punishment. In the early 1700s, imprisonment was practically not existent, and the primary sanctions were transportation to the Americas and execution; in fact, there were more than 200 capital offenses by 1800, a period known as *The Bloody Code*. The British penal system was put into crisis when the American Revolution abruptly eliminated the Americas as a penal colony in 1776. This led to the first, albeit temporary, mass use of prison sentences; transportation did not resume until the establishment of a penal colony in Australia. However, by the end of the nineteenth century, capital punishment had been abolished for most offenses by a series of offense-specific acts in the mid-1800s, transportation had been (mostly) abolished in 1853, and the modern-day prison sentence was the primary form of punishment.

The accounts of more than 200,000 criminal cases tried at the Old Bailey Criminal Court in London between 1715 and 1900 were published in *The Proceedings of the Old Bailey*, which has been digitized and published by *The Old Bailey Proceedings Online* (Hitchcock et al. 2012).<sup>7</sup> Our empirical analysis focuses on almost 100 years of this data (1772–1871). From this remarkable dataset, we extracted information identifying the unique case, session date, defendant’s name, gender and age, charged

<sup>6</sup> See a 2016 *Washington Post* article for historical examples: <https://www.washingtonpost.com/news/in-theory/wp/2016/04/08/history-is-clear-juries-were-supposed-to-be-able-to-overturn-laws/>.

<sup>7</sup> Vickers and Ziebarth (2016) uses these data (1835–1913) to study changing demographics in crime; see Bindler and Hjalmarsson (2017b) for an extension to earlier years. Economic historians (Voth 1998; Horrell, Humphries, and Sneath 2015; and Kelly and Ó Gráda 2016) use the proceedings to construct time-use budgets and measure the value of property.

offense, as well as broad and detailed verdict and sentencing outcomes. We also manually coded judge and jury names from 1750 to 1822 and criminal history after 1832.

Given the context, it is natural to question the relevance of our study to today. Though transportation “beyond the seas” clearly no longer exists, capital punishment is still used and actively debated in many countries, including the United States today. Potential jurors in a US capital case can even be dismissed for cause if they oppose the death penalty due to the implicit assumption that such an individual cannot be impartial—an assumption that we empirically test in this paper. This disqualifies a large proportion of jurors today: a recent survey of California jurors today found that 24 percent of respondents would not be comfortable convicting a murderer knowing that a death sentence would follow while 32 percent would automatically vote for life imprisonment (Garrett, Krauss, and Scurich 2017). Even though juries typically decide sentences in capital cases today (as opposed to judges in our context), this was not mandated until a 2002 Supreme Court decision that affected 13 states (Iyengar 2011). Moreover, if one finds a jury incapable of being impartial in the adjudication stage of a trial studied in our paper, when they are not directly imposing a death sentence themselves, it seems likely that such partiality would carry over to the sentencing phase.<sup>8</sup> In addition, one can draw parallels to other policies where lawmakers and society (jurors) might diverge in their perceptions of the proportionate harshness of the sanction. This may apply, for instance, to life without the chance of parole, especially for juveniles, and mandatory minimums for drug sentences.

There are two key advantages to studying the abolition of capital punishment in this historical context. First, it provides a large and unambiguous decrease in punishment severity, which simply cannot be observed today. Second, the differential timing in the abolition of capital punishment across offenses allows for a difference-in-difference design to retrieve the (causal) effect of changes in punishment severity on jury verdicts in a single jurisdiction. Specifically, for each of the 26 offense categories in our data, we identify whether the offense was never, always, or once capital eligible and in the latter case the year that capital punishment was abolished, which ranges from 1813 (fraud) to 1856 (arson). Intuitively, our research design compares the change in the chance of conviction in the years surrounding the abolition of capital punishment for “treated” offenses—i.e., those for which the capital status changed—to that for “control” offenses—i.e., those which were never or always capital eligible. Such a design controls for other changes occurring during this period in both the criminal justice system (e.g., the introduction of the Metropolitan Police) and society more generally (e.g., the Industrial Revolution or more general changes in social norms).

<sup>8</sup> Another question regarding external validity relates to the limited (and declining) use of jury trials today. Despite a defendant’s sixth amendment right to a trial by their peers, more than 95 percent of felony convictions result from a guilty plea (BJS, <https://www.bjs.gov/index.cfm?ty=tp&tid=23>, accessed September 11, 2017). How are historical jury decisions relevant to this modern context? While jury trials are relatively low in number today, cases that go to trial tend to be relatively high stakes (as in our historical context) and often have death and life without parole on the table. According to Texas’s Annual Statistical Report (Texas Judicial Branch 2014), around 50 percent of convictions for both capital murder and murder are decided by juries.

Our empirical analyses find that the decrease in expected punishment arising from the abolition of the death penalty significantly increased the chance of any conviction by 7.6 percentage points (10.6 percent) and the chance of conviction of the original charge by 16 percentage points. The former is driven by violent and sex offenses as well as fraud cases while the latter is seen for all offenses, and especially for property offenses. This was accompanied by a large decrease in the chance of a recommendation for mercy for violent and fraud offenses—as mercy was no longer needed to spare someone death. Thus, before the abolition of capital punishment, the jury used two actions to spare defendants the chance of a death sentence: for property crimes, they convicted defendants of a lesser charge that was not capital eligible while for violent and fraud offenses, they either acquitted or convicted with a recommendation for mercy.

Heterogeneity analyses indicate that a jury's reluctance to convict on a capital charge is unequal across defendants: juries were significantly more reluctant to convict females than males of a (violent) capital offense. Though less precise, we also find suggestive evidence of a similar reluctance to convict first-time offenders than repeat offenders.

A causal interpretation of these results relies on the assumptions of parallel trends and that the timing of the offense-specific abolition was random with no anticipatory effects. In support of these assumptions, we demonstrate the robustness of our results to offense-specific time trends. In addition, we find no significant “lead” effects in event-study specifications. An additional implicit assumption is that the composition of cases and corresponding quality of evidence presented to the jury did not change with the reforms. This could occur, for instance, if a change in offender criminal behavior or defendant plea behavior impacts the types of cases presented to the jury. In other words, an increase in the quality of evidence could feasibly yield the same pattern of results (i.e., an increase in conviction rates) and confound the effect of interest. However, we provide direct empirical evidence that the results are robust to including all pleas as guilty of the original as well as guilty of a lesser offense jury verdicts; there is no significant increase in the chance of pleading guilty; there is no observable change in criminal behavior; and there is no significant increase in the quality of evidence, measured using keyword searches for police, evidence, and witness in the proceedings.

Since almost all offense categories were “treated” contemporaneously with the temporary halt of transportation, i.e., almost all offenses were transportation eligible, we are unfortunately limited to using the simple pre-post research designs in this context. As a result of the American Revolution, the share of sentences to transportation decreased from 75 percent to 0 percent in 1776 and resulted in an increase in sentences to prison and manual labor in the hulks of ships. Notably, this sharp and unexpected change in expected punishment is exogenous to the criminal justice system. In particular, defendants charged with transportation-eligible *noncapital* offenses faced an unambiguous decrease in punishment severity during the war, as imprisonment was substituted for transportation. The temporary halt of transportation increased the chance of conviction by about 5 percentage points for these individuals. The obvious weakness in the simple pre-post identification strategy—namely the inability to conclusively separate the effect of the reform



from other things changing with the war—limits the causal interpretation of the transportation results. Nevertheless, we believe this “experiment” adds to a complete picture of the role of punishment severity in jury verdicts.

This paper provides empirical evidence that punishment severity may impact the ability of a jury to be impartial. We cannot distinguish whether the lower pre-reform conviction rates are attributable to jury nullification due to disproportionately harsh sentences (i.e., a jury’s willingness to allow for “Type II errors” or false acquittals) or a higher “beyond” a reasonable doubt threshold when the consequences of a “Type I error” or false conviction are greater. However, the fact that abolishing capital punishment has such large impacts on behavior during a period when capital punishment still had a high acceptance rate in society is striking. It is certainly suggestive that the chance of a death sentence may significantly impact jury behavior today, when the death penalty is much less socially accepted. It also suggests that the behavior of jurors who are not fundamentally opposed to the death penalty may still be impacted by the potential sanction. Further, this lack of impartiality may be applied unequally across defendants, leading to undesired biases in jury decision making.

The remainder of the paper proceeds as follows. Section I provides institutional details on the criminal justice system and changing sentencing regimes in the eighteenth and nineteenth centuries. Section II describes the data and defines the treatment and control groups for each experiment. Sections III and IV present the results concerning the abolition of capital punishment and halt of transportation, respectively. Section V concludes.

## I. Institutional Background

### A. *The Rise and Fall of Capital Punishment, Transportation, and Incarceration*

The years from 1715 to 1900 in England represent a period of dynamic change in the criminal laws governing sentencing, providing a unique natural experiment to study how changing expected punishment affects the behavior of various agents in the criminal justice system. This section provides a broad overview of this history and is based on *The Old Bailey Proceedings Online*, original acts obtained from the Parliamentary Archives, and a number of books summarizing these acts (Cook and Keith 1975; Hitchcock and Shoemaker 2015).

In 1688, there were approximately 50 capital offenses. The number of offenses classified as capital began to rise with the Waltham Black Act of 1723, which introduced the death penalty for over 50 more offenses. Numerous parliamentary acts, largely motivated by a desire to protect the property of land-owning classes, subsequently increased the number of capital offenses to 160 in 1765 and more than 200 in the early 1800s, including offense subcategories. This period became known as *The Bloody Code* because of the high number of capital offenses and the public and/or bloody spectacle made of executions. At the turn of the nineteenth century, even crimes viewed today as petty crimes (e.g., pickpocketing and shoplifting) were capital.

A movement to reform the criminal justice system, led by Sir Robert Peel, began in the 1820s with the passage of the Judgment of Death Act of 1823. This act made

the death penalty discretionary for almost all then capital crimes except murder and treason. Though judges still had to officially enter a death sentence (as seen in the data), this sentence could later be reduced at the judge's discretion. Additional acts reduced the number of offenses even eligible for the death penalty throughout the early and mid-1800s.<sup>9</sup> The public spectacle of executions ended in 1868. Figure 1 demonstrates that more than 75 percent of the 26 offenses in our analysis were capital eligible from 1715 to 1820. This then sharply decreased to about 15 percent, held steady until the early 1860s, and then sharply decreased again. Online Appendix Table 2 lists the offenses underlying this figure and the titles of the corresponding acts that abolished capital punishment.

In the early 1700s, a not insignificant share of offenders could escape capital punishment by invoking the "benefit of clergy."<sup>10</sup> Transportation provided an alternative that individuals could not escape on these grounds. The first Transportation Act (1718) allowed individuals convicted of a clergyable offense to be transported to America for seven years; returning from transportation was a capital offense. Transportation was unexpectedly halted in 1776 due to the American Revolution. Faced with a penal crisis, the *Hulks Act* (1776) was passed, allowing male convicts to be put to hard labor (dredging the river Thames) and held in the hulks of ships. Poor conditions on the hulks (with frequent escape attempts and high risk of death due to poor nutrition and illness) and growing resentment towards the overcrowded prison system contributed to the eruption of *The Gordon Riots* on June 2, 1780 (Hitchcock and Shoemaker 2015). During the weeklong riots, many prisons were attacked, and prisoners escaped or were released. Military intervention ended the riots, leaving a temporary military presence and a distrust of the lower classes by "respectable Londoners" (Hitchcock and Shoemaker 2015).

To combat the growing unrest among the people, the courts resumed transportation sentences in October 1781, despite the lack of a viable new penal colony; those receiving such sentences were imprisoned. A new penal colony was finally established in 1786 in Botany Bay Australia, to which the First Fleet (eleven ships with more than 700 convicts) set sail in May 1787.<sup>11</sup> Being "transported for life beyond the Seas" to Australia was seen as a worse punishment than transportation to the Americas. The voyage commonly took four to six months, during which time many became ill or died. Upon arrival, the convicts were often put to hard labor in gangs developing infrastructure. Discipline was harsh—lashes, chain gangs, or being sent to the most remote penal colonies in Australia. Transportation rose throughout the 1820s and 1830s, as it replaced capital punishment as the maximum sentence. Transportation was abolished through the Penal Servitude Acts of 1853 and 1857,

<sup>9</sup> The death penalty was abolished for the remaining offenses in 1965 (murder), 1971 (arson on the docks), 1981 (espionage), and 1998 (high treason).

<sup>10</sup> Since the Middle Ages, a criminal could be handed over to his church for clergyable offenses. To prevent too many criminals from getting off, many offenses were reclassified as clergyable: murder, rape, robbery, burglary, and pickpocketing in the 1500s; and housebreaking, theft from a dwelling, shoplifting (of more than 40 and 5 shillings, respectively), and sheep/cattle theft in the 1600-1700s (Beattie 1986).

<sup>11</sup> Four "experiment" ships sent to Africa, America, and Honduras in 1782 and 1785 failed due to the "mutinous spirit of the convicts" and "rejection by the destination populations" (Hitchcock and Shoemaker 2015).





FIGURE 1. SHARE OF CAPITAL-ELIGIBLE OFFENSES (1715–1900)

*Notes:* The figure shows the share of offense categories in the sample that are eligible for capital punishment between 1715 and 1900. These shares are based on the 26 offense categories included in our analysis: animal theft, mail, stealing from master, theft from place, shoplifting, larceny, arson, burglary, housebreaking, receiving, manslaughter, murder, assault, wounding, robbery, rape, sexual assault, sodomy, coining offenses, embezzlement, forgery, fraud, bigamy, libel, perjury, and perverting justice.

*Source:* *The Old Bailey Proceedings Online* and own calculations

with an increased perception that it was inhumane and did not deter.<sup>12</sup> The former replaced transportation for seven years with four-years penal servitude, retaining transportation for only long-term cases. The 1857 Act abolished transportation for these cases, though the last convict ship did not set sail until 1867.

The idea of imprisonment as a mainstream sentencing model dates back to the American Revolution, when a substitute was needed for transportation. Newgate, the main London prison in the 1700s, was largely used to hold individuals awaiting trial or execution. With the previously mentioned reforms, the use of imprisonment became the primary sanction. Millbank Penitentiary opened in 1821, with 860 cells. Pentonville (a model for many future prisons) opened in 1842, with capacity for 520 prisoners to spend up to 18 months in solitary confinement.

Our identification strategy capitalizes on the sharp changes in punishment severity resulting mainly from two natural experiments in this historical context. Figure 2 illustrates these by presenting the share of convicted offenders at the Old Bailey (as recorded in the proceedings) sentenced to death (black line), transportation (dark grey line), and imprisonment (light grey line). The first experiment focuses on the decrease in the share of death sentences from around 25 percent to almost zero in the mid-1800s due to the abolition of capital punishment. Second, the temporary

<sup>12</sup> See Emsley, Hitchcock, and Shoemaker (2011) on punishments at the Old Bailey: <http://www.oldbaileyonline.org/static/Punishment.jsp#transportation>.

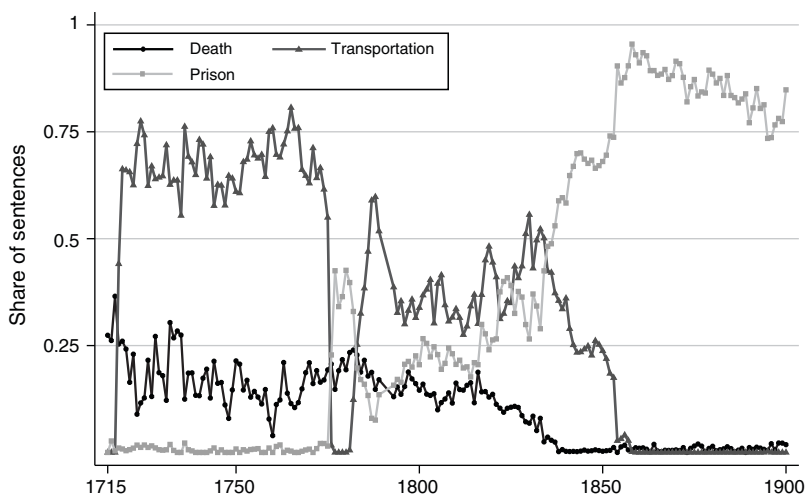


FIGURE 2. SHARE OF SENTENCES—DEATH, TRANSPORTATION, AND PRISON (1715–1900)

*Notes:* The figure shows the annual share of convicted cases that result in a death penalty (black line), penal transportation (dark grey line), and prison (light grey line) in the sample between 1715 and 1900. These shares do not sum to 100 percent since some other, less common sentence categories (such as corporal punishment or miscellaneous punishment) are not shown.

*Source:* *The Old Bailey Proceedings Online* and own calculations

halt of transportation during the American Revolution results in a drop in transportation sentences from around 75 percent in the first half of the 1700s to 0 percent during the war, with a corresponding temporary increase in prison sentences.<sup>13</sup> The abolition of transportation is seen in the 1850s and is briefly considered as a third natural experiment in Section IV.

### B. London and the English Jury System in the Eighteenth and Nineteenth Century

The data come from trials at the Old Bailey, which was the central criminal court for the City of London and the surrounding County of Middlesex; it was responsible for the most serious criminal trials, including all felonies. The Old Bailey catchment area, however, changed a bit over time, with an 1830s expansion to Essex and because other courts trialed less serious crimes. Criminal cases were tried by a jury upon a grand jury's decision of sufficient evidence to proceed. The legal system at the beginning of this period gave little attention to defendant rights. This is anecdotally supported by shockingly short trials (timed by an Old Bailey chaplain to be, on average, eight minutes in the early nineteenth century; Feeley 1997). Little evidence was presented by the defendant or attorneys (who did not yet play a significant role); rather, Feeley (1997, 190) describes the defendant's testimony to be more consistent

<sup>13</sup> The high prevalence of transportation is driven by the most prevalent offense—larceny, which is classified as noncapital. The share of prison and death sentences do not completely offset the decrease in transportation; other sentences (not shown), especially corporal punishment, were also used increasingly during the Revolution.

with the modern-day sentencing phase of a trial and consists “of mixtures of weak denial, pleas for mercy, and testimony of good character.” This began to change when the burden of proof shifted to the prosecution with the presumption of innocence (1827) and felony indictments were entitled to defense attorneys (1836). Jury deliberations also changed in ways that likely increased the chances of a fair trial; in particular, until 1858, juries were not allowed fire, food, or drink until a verdict was reached.<sup>14</sup>

For most of the period studied in this paper, the jury selection process was governed by the Juries Act of 1825, which defined men between ages 21 and 60, who resided in England and had land/wealth of an appropriate threshold as eligible for jury service.<sup>15</sup> To be geographically representative, separate juries were selected for London and Middlesex cases. Individuals in the jury pool were selected from an annually updated master list and received a summons ten days prior to the beginning of each session. Though little is known about how the pool was selected (Langbein 1987), the 1825 Act does detail how to seat a jury of the first 12 randomly drawn men (from a box with parchment cards of equal size) not struck for cause (including ineligibility to be in the pool in the first place).

A unique feature of this period is that jurors had extensive experience. In contrast to the US jury system today, the same jury decided one case after another for the entire session, which typically lasted for a few days. In addition, almost every jury had at least one juror who had served previously on a jury in another session. Having coded the juror and jury names for more than 50 years of Old Bailey trials, we can see this explicitly in the data. From 1751–1808, there were on average 42 trials per jury, and more than 90 percent of Middlesex and 75 percent of London juries had at least one experienced juror, oftentimes many more.<sup>16</sup> Therefore, a reasonable assumption is that the jury knew (or had an expectation of) the punishment associated with a guilty verdict for various offenses. This is even more likely to be true from the 1840s on, when the judge announced the sentence immediately after the verdict; that is, the jury observed the sentence for each case before hearing the next (Bentley 1998). Prior to the 1840s, sentences were given to all convicted defendants on the last day of the hearings/session. Throughout the analysis period, however, jurors likely formed expectations about sentencing by regularly reading the proceedings themselves, which were published for public consumption.

Not surprisingly, other aspects of the criminal justice system also changed during this period. Perhaps the most notable institutional change is the introduction of the Metropolitan Police in 1829; it consisted of about 3,000 uniformed men tasked with patrolling the streets to deter crime. The late 1700s and early 1800s in England were characterized by the industrial revolution that led to agglomeration and urbanization. London’s population increased from around 630,000 in 1715 to over one

<sup>14</sup> Source: Old Bailey website.

<sup>15</sup> See Bentley (1998) for a summary of the Juries Act of 1825 and the English criminal justice system in the 1800s.

<sup>16</sup> See Bindler and Hjalmarsson (2017a), testing for path dependency within these consecutive jury decisions, for additional details and statistics about the jurors and their experience.

million in 1801, the year of the first census, and again tripled by 1860.<sup>17</sup> By 1815, London was the largest city in the world.

## II. Data

### A. *The Proceedings of the Old Bailey*

*The Proceedings of the Old Bailey* were published from 1674 to 1913, though cases were not consistently recorded until 1715. After each monthly session, the proceedings published an account of the criminal cases trialed at the Old Bailey, though the details varied over time and across cases. As described on the Old Bailey Online, the proceedings initially provided entertainment for the population, with detailed transcripts of the most colorful cases. By 1787, the proceedings had a quasi-official status, as the City of London had to subsidize the publishers and, from 1778, “demanded that the proceedings should provide a ‘true, fair, and perfect narrative’ of all the trials,” leading to approximately equal coverage of all trials.

The records have been digitized by *The Old Bailey Proceedings Online*.<sup>18</sup> We obtained the proceedings for each of the 2,000 court sessions in XML files and extracted information identifying the unique case, session date, defendant’s name, gender and age, offense category, and broad and detailed verdict and sentencing outcomes. Though the text of the proceedings describes all charges, only the main (most serious) charge is tagged in the proceedings online. That is, a defendant may be charged with multiple offenses, but we only observe the most serious offense. We term this offense the original charge. The broad verdict data indicate whether the jury found the defendant guilty, while the detailed data help distinguish between guilt of the original charge or a lesser offense than originally charged and whether he/she pled guilty. For instance, a defendant charged with murder may be found guilty of manslaughter. The broad punishment variable indicates the primary sentence issued by the judge—death, transportation, imprisonment, corporal punishment, miscellaneous, or no punishment.

The proceedings also contain judge, jury, and juror names for most cases between 1750 and 1822 that are not easily extracted with the XML tags. In particular, the front matter of the proceedings lists the names of the jury (e.g., the first London or second Middlesex jury) and each of the twelve jurors on the jury. While these names are tagged, the tags do not link to the cases seen by each jury; we manually transcribed this information. Each session has at least one Middlesex and one London jury, allowing us to use the jury name to control for jurisdiction.

Finally, we manually transcribed information on the criminal history of the offender, which is available from the 1830s onwards and contains information on whether the defendant had been in custody once before (from 1832), more than once (from 1839), or whether they were known associates of bad character (from 1835).

<sup>17</sup> <https://www.oldbaileyonline.org/static/Population-history-of-london.jsp> (accessed September 14, 2016).

<sup>18</sup> <http://www.oldbaileyonline.org/>, maintained by HRI Online Publications, provides a tremendous amount of information about the history of the proceedings, the digitization process, as well as a search engine.

Previously, criminal history was largely irrelevant, since most known criminals were sentenced to death or transported.

As some cases have multiple defendants, the final dataset is created at the case by defendant level; each observation refers to a unique defendant. From 1715 to 1900, there are 217,939 defendant-case observations. We exclude the 2,057 observations from 1790 to 1792, when the proceedings selectively reported only guilty verdicts, and 751 observations with obvious misreporting or missing values in crucial variables. The raw data provide a high level of detail with respect to the offense. As indicated in online Appendix Table 1, we classify offenses into broad categories: property, violent, sex, fraud, and other. We omit less than an additional 5 percent of observations over this 200-year period that simply cannot be analyzed in any meaningful way.<sup>19</sup> These restrictions result in an analysis sample of 206,733 defendant-case observations from 1715 to 1900, which we break down into the experiment specific estimation samples. Online Appendix Figure 1 displays the annual number of cases in each broad offense category.

### B. Coding Treatment Offenses and Years for Each Experiment

A crucial step in our analysis is coding the treated offense categories and treatment years for each experiment. We do so using a two-step approach. First, we identify observable discontinuities in the share of death and transportation sentences in our data. Second, we compare the timing of the observed discontinuity to that of the historical events or changes in laws, obtained from historical sources whenever possible. That is, we primarily use the observed discontinuities to identify the years in which we should focus our search for the actual law. We follow this procedure because the long time horizon (200 years) and complicated nature of these historical laws make it practically impossible to track and find all offense and sub-offense-specific relevant laws. See online Appendix Figure 2 for two examples of original laws.

More specifically, for the natural experiment concerning the halt of transportation, we identify the set of transportation-eligible offenses (those with a positive share of transportation sentences) immediately prior to the Revolution in 1776. Only these offenses were actually *treated* by this event. We drop untreated offenses since the small sample size (780 cases from 1772 to 1789) does not allow for a meaningful control group to implement a difference-in-difference design. Table 1 indicates those offenses dropped in the sample cleaning process (D1), those dropped because they are untreated offenses (D2), and those treated (T).

For the death penalty natural experiment, we capitalize on the offense-specific variation by coding a unique treatment period for each offense; we code the first

<sup>19</sup> We exclude: (i) 2,649 cases of offenses for which the overall number of trials is too low to analyze, (ii) 865 cases with offenses containing unusually many defendants (conspiracy and riot), (iii) 186 cases for offenses redefined during the period and for which the redefinition cannot be clearly distinguished from changes in sanctions (kidnapping), and (iv) 4,698 cases with no specific offense given in the data. Specific offense categories and the associated number of cases from 1715–1900 dropped in (i) are bankruptcy (404), barratry (4), concealing a birth (474), extortion (323), game law offenses (47), illegal abortion (90), infanticide (328), keeping a brothel (88), petty treason (14), piracy (7), religious offenses (17), return from transportation (378), seditious libel (45), seditious words (35), seducing allegiance (20), tax offenses (189), threatening behavior (145), treason (39), and vagabond (2).

TABLE 1—TREATMENT/CONTROL OFFENSES AND TREATMENT YEAR FOR EACH EXPERIMENT

Offense	Capital punishment			Transportation	
	Offense-specific laws			American Revolution	
	+/- 10 years around treatment year (T)			1772–1789	
	Treatment	Year	Cases	Treatment	Cases
<i>Property offenses</i>					
Animal theft	T	1832	1,168	T	435
Arson	T	1856	111	D1	19
Burglary	T	1837	1,081	T	1,323
Housebreaking	T	1833	1,396	T	164
Larceny	C (never)	NA	32,278	T	8,181
Mail	T	1834	74	D1	5
Receiving	T	1837	3,567	T	686
Shoplifting	T	1820	763	T	441
Stealing from master	C (never)	NA	4,696	D1	0
Theft from place	T	1832	3,706	T	1,537
<i>Violent and sex offenses</i>					
Assault	C (never)	NA	185	D1	5
Manslaughter	C (never)	NA	295	D2	14
Murder	C (always)	NA	222	D2	161
Robbery	T	1837	859	T	1,529
Rape	T	1841	228	D2	63
Sexual assault	D	NA	0	D1	0
Sodomy	T	1832 (1860)	81	D2	15
Wounding	T	1837 (1861)	825	D2	35
<i>Fraud offenses</i>					
Coining offenses	T	1832	893	D2	337
Embezzlement	C (never)	NA	1,650	D1	3
Forgery	T	1832	581	D2	155
Fraud	T	1813	160	T	151
<i>Other offenses</i>					
Bigamy	C (never)	NA	225	D1	20
Libel	C (never)	NA	23	D1	1
Perjury	C (never)	NA	102	T	100
Perverting justice	T	1831	83	T	77
<i>Total</i>	<b>26</b>		<b>55,252</b>	<b>26</b>	<b>15,457</b>
Treatment	16		15,576	11	14,624
Control/untreated	9	<b>Median: 1833</b>	39,676	7	780
Dropped	1		0	8	53

*Notes:* The table shows the classification of offenses into treatment (T) and control (C) groups for the capital punishment sample as well as into treated (T) and untreated (D2) offenses for the transportation sample, the assigned treatment year, and the number of observations for each of the analyzed natural experiments. D and D1 indicate which offenses were dropped for the capital punishment and transportation “experiments,” respectively, generally because of insufficient sample size.

*Source:* *The Old Bailey Proceedings Online*, various sources as specified in the text (laws), and own calculations

treatment year as the year when the share of death sentences drops to zero for that offense.<sup>20</sup> Offenses with no such discontinuity are classified as always or never capital and assigned to the control group. Figure 3 provides examples—murder

<sup>20</sup> Thirty-seven cases appeared to be anomalies, with death sentences occurring after the drop to zero. A close reading of the proceedings made it clear that these cases had multiple related charges—one for which capital punishment had been abolished (e.g., burglary) and the other still capital (e.g., felonious wounding). We recoded such cases as the appropriate more severe offense.



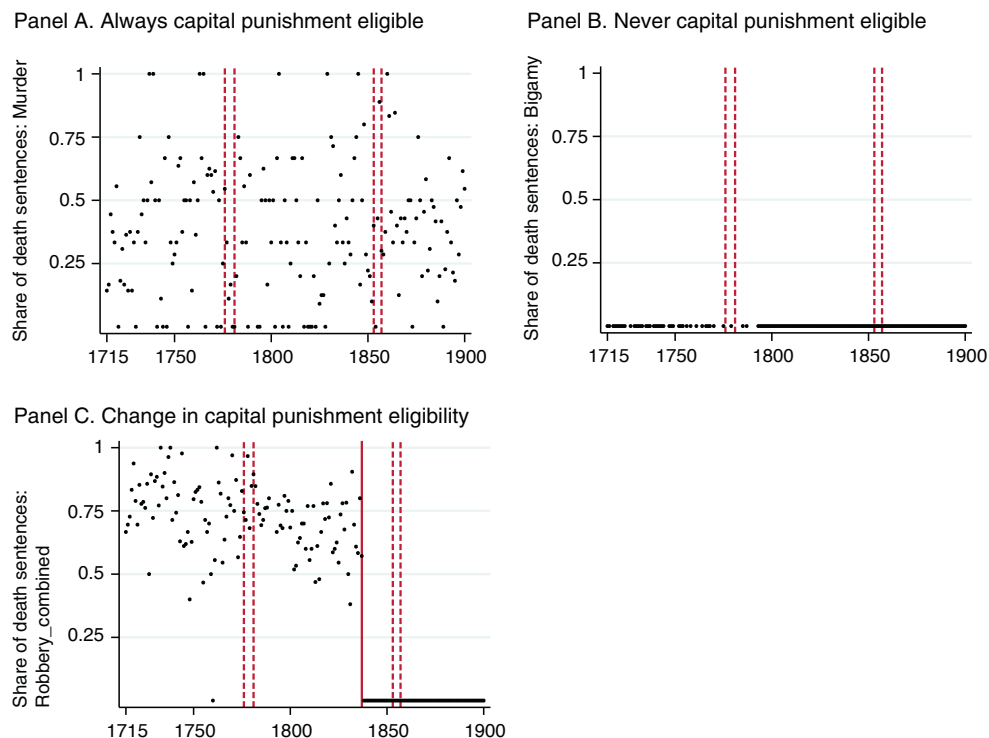


FIGURE 3. IDENTIFYING THE TIME OF TREATMENT—ABOLITION OF CAPITAL PUNISHMENT

*Notes:* The figure shows the annual share of convicted cases in the sample that were sentenced to death for murder (panel A, always capital punishable), bigamy (panel B, never capital punishable), and robbery (panel C, change in capital punishment eligibility). The dashed vertical lines mark the years that were affected by changes in penal transportation (American Revolution and abolition of transportation); the solid line in panel C marks the year of treatment, i.e., the first year in which the observed share of capital punishment is equal to zero.

*Source:* *The Old Bailey Proceedings Online* and own calculations

(always capital), bigamy (never capital), and robbery (reformed in 1837)—of the types of discontinuities in the data. Each graph shows the share of death sentences over time; the solid line marks the discontinuity year.<sup>21</sup> See online Appendix Table 2 for the corresponding offense-specific historical acts from the House of Lords Parliamentary Archives and additional online sources. Table 1 indicates treatment and control offenses for the capital punishment experiment and the first years of treatment. For all but three offenses (receiving, fraud, and perverting justice), we identified the actual law pertaining to the abolition of the death penalty, the dates of which correspond exactly to the years observed in the data. For the remaining offenses, we rely solely on the discontinuity in the data; the results are robust to excluding these offenses. Though there are law changes in both 1837 and 1856 for arson, there is a discontinuity only for the latter; we thus code 1856 as the reform year.

<sup>21</sup> Such figures are available for each offense upon request for each experiment.

While one would ideally like to code the exact month of the reform, this is unfortunately not feasible for two reasons. First, the date cannot be reliably backed out from examining discontinuities in the data since zero sentences in a given month can also arise due to there being zero charges in an offense category, i.e., the monthly concentration of each offense is not high enough to look at the raw data. Second, while some laws explicitly state the reform date, it is unclear if it was put into practice immediately; indeed, looking at the monthly sentencing data highlights at least two instances (burglary and wounding) where the law abolishing the death penalty is written in July but sentences are seen through October. Our coding of the reform as the first year with zero death sentences therefore likely assigns some “treated” cases to the pre-reform period. We believe this to be of minimal concern, however, since the results are completely robust (and available upon request) to excluding from the analysis the reform year, the first pre-reform year, and the first post-reform year.

It is important to highlight that our two-stage approach to identifying treatment and control offenses capitalizes on discontinuities observable in sentencing variables. That is, we do not look for discontinuities in the outcome (verdicts), but rather in a measure of the “treatment.” One offense category for which this may be a concern, however, is larceny. As seen in online Appendix Table 1, we combine the offenses of grand larceny (theft of more than 1 shilling), petty larceny (theft of less than 1 shilling), simple larceny, and pocket picking into a single larceny offense.<sup>22</sup> Capital punishment did exist for larceny if the goods were of a value greater than a specific threshold, which changed over time. In practice, however, death sentences are almost never seen for larceny because the juries had the ability to convict the defendant of a lesser (noncapital) charge. For this reason, we demonstrate that both our overall results and property crime results are robust to excluding larceny.

Table 1 indicates the number of observations for each treated and untreated or dropped offense from 1772 to 1789 for the transportation experiment and, for descriptive purposes,  $\pm 10$  years around the offense-specific year that capital punishment was abolished. For offenses in the capital punishment control group, we report the number of observations in a window around the median reform year of 1833. A number of facts stand out. First, dropped (but treated) offenses are those that cannot be reliably studied given either the rarity with which they are observed or that the offense-specific abolition of capital punishment falls into the same time window as the abolition of transportation (wounding). Second, contrary to the transportation sample, there are 16 offenses for which capital punishment was abolished (15,576 observations) and nine control offenses (39,676 observations). There is substantial variation in the abolition year; the earliest year is 1813 for fraud while the latest is 1856 for arson.<sup>23</sup>

<sup>22</sup> This allows for a continuously defined “larceny” variable over the entire period despite changing definitions.

<sup>23</sup> We also point out that our classification of capital eligibility is based on the session date and not the offense date, which is not systematically available in the data. We believe this to be unproblematic given the generally short time between offense and trial and since at least some laws were defined on the basis of the conviction date (see the example in the online Appendix). An examination of the text of the proceedings supports this. For instance, for the February 16, 1832 proceeding, almost every offense occurred in the six weeks leading up to trial.

### *C. Summary Statistics*

Table 2 presents summary statistics for the whole sample (1715–1900) as well as the subsamples corresponding to each experiment: 1803–1871 for capital punishment and 1772–1789 for transportation. From 1715 to 1900, there were 1,748 sessions at the Old Bailey, more than 900 of which are included in our analysis periods. In terms of the broad offense groups, 73 percent of cases are property offenses while the remaining are classified as violent (10.1 percent), sex (1.8 percent), fraud (13.3 percent), and other (2 percent); however, property offenses comprise almost 88 percent of all cases during the Revolution. In addition, each category is represented in both the treatment and control groups for the death penalty experiment, though the control group consists of a larger share of property offenses and the treatment group relatively more violent and fraud offenses.<sup>24</sup>

The average recorded defendant age from 1715–1900 is 27.6. We do not control for age, however, since it is missing for 99 percent of cases in the transportation experiment and only reported for guilty defendants in later years. More than 21 percent of defendants are female, with a larger share during the transportation period (27 percent). In fact, from the early 1700s until 1900, the share of female defendants is trending down towards contemporary levels: 36 percent of defendants were female prior to 1750, compared to 28 percent from 1750–1800, 22 percent from 1800–1850, and 13 percent from 1850–1900. Finally, slightly more than 10 percent of defendants in both the treatment and control groups for the death penalty experiment have some criminal history (when such data was recorded). From 1750 to 1822, 24 percent of the juries were for London (versus Middlesex) cases.

There are four main outcomes for a defendant at trial. The first is a choice of whether to plead guilty. It is clear from Table 2 that the practice of pleading guilty increased over time—from 0.2 percent of cases in the 1700s to more than 14 percent from 1803 to 1871. These statistics are largely driven by pleas for forgery and fraud, a concern we return to later. If a defendant does not plead, then there are three possible jury verdicts: convict of the original charge, convict of a lesser charge, or acquit. For cases decided by the jury, 58 percent of defendants from 1715–1900 were convicted of the original charge and almost 10 percent of a lesser charge; about 32 percent were acquitted (42 percent during the Revolution). Breaking up capital punishment cases into treated and control offenses shows that only 57 percent of treated offenses are convicted of their original charge compared to 73 percent for controls while they are more likely to be found guilty of a lesser offense (11 percent versus 1 percent) or acquitted (32 percent versus 26 percent). Finally, if the jury convicts, it can recommend mercy; this is the only jury input into sentencing. Conditional on conviction, mercy is recommended in about 6 percent of cases from 1715–1900, with a larger share during the capital punishment period. Mercy recommendations were concentrated amongst property offenses (with the largest shares for larceny, stealing from master, and theft from place) but seen for all offenses.

<sup>24</sup> All offenses classified as capital have a substantial share of pre-reform death sentences—the lowest (around 30 percent) is for shoplifting; thus, there are no capital offenses available to use, for instance, as a pseudo-control offense.

TABLE 2—SUMMARY STATISTICS

Variable	All	Capital punishment		Transportation
	1715–1900	Treatment 1803–1871	Control 1803–1871	American Revolution 1772–1789
<i>Sample</i>				
Number of observations (N)	206,733	49,285	76,673	14,624
Number of sessions (N)	1,748	703	703	153
Average number of cases per session	150.0	172.8	211.1	97.64
Average number of defendants per case	1.483	1.762	1.265	1.512
<i>Offenses</i>				
Property offense (0/1)	0.729	0.577	0.893	0.873
Violent offense (0/1)	0.101	0.137	0.032	0.105
Sex offense (0/1)	0.018	0.020	0.006	
Fraud offense (0/1)	0.133	0.262	0.051	0.010
Other offense (0/1)	0.020	0.005	0.018	0.012
<i>Defendants</i>				
Average age	27.57	27.10	26.40	
Age missing (0/1)	0.376	0.237	0.230	0.985
Male (0/1)	0.786	0.813	0.781	0.728
Any known criminal history, from 1832 (0/1)	0.111	0.106	0.101	
<i>Juries and judges</i>				
Average number of juries per session	3.072			3.245
London jury (0/1)	0.243			0.277
Number of judges	104			30
<i>Pleads and Verdicts</i>				
Plead guilty (0/1)	0.136	0.140	0.148	0.002
Guilty of original offense by jury (0/1)	0.583	0.572	0.731	0.453
Guilty of lesser offense by jury (0/1)	0.096	0.113	0.010	0.139
Acquitted by jury (0/1)	0.324	0.318	0.259	0.416
Recommended for mercy (0/1, conditional on conviction)	0.061	0.071	0.093	0.030
<i>Sentences conditional on guilty by jury or plea</i>				
Capital punishment (0/1)	0.069	0.123	0.004	0.187
Transportation (0/1)	0.294	0.259	0.308	0.381
Imprisonment (0/1)	0.522	0.564	0.578	0.186
Corporal punishment (0/1)	0.042	0.011	0.038	0.179
Miscellaneous punishment (0/1)	0.045	0.020	0.042	0.055
No punishment (0/1)	0.030	0.022	0.030	0.012

Notes: The table shows summary statistics for the variables in the analysis sample for each of the analysis periods. Note that guilty of original offense, guilty of lesser offense, and acquitted by jury are defined for just the sample with a jury verdict. Where not otherwise specified, the mean of the variable is shown. The treatment group contains those offenses for which capital punishment is abolished and the control group those that are never or always capital eligible (see Table 1 for details).

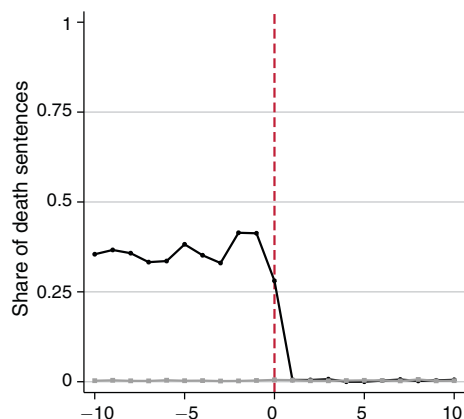
Source: *The Old Bailey Proceedings Online* and own calculations

III. The Impact of Abolishing Capital Punishment on Jury Decision Making

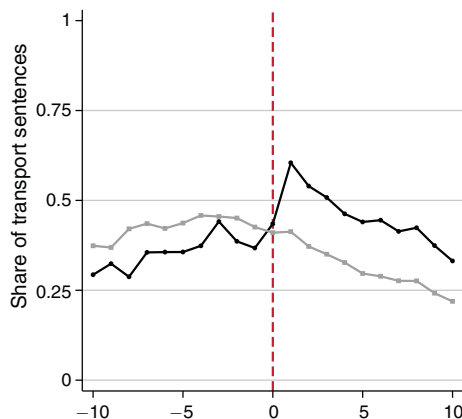
A. Graphical Evidence of the Treatment—Capital Punishment

We begin by demonstrating the impact of abolishing capital punishment on sentences to death, transportation, and prison for both the treatment and control groups in Figure 4. The figure shows the share of each sentence in the ten years before and

Panel A. Treatment and control group, death penalty



Panel B. Treatment and control group, transportation



Panel C. Treatment and control group, prison

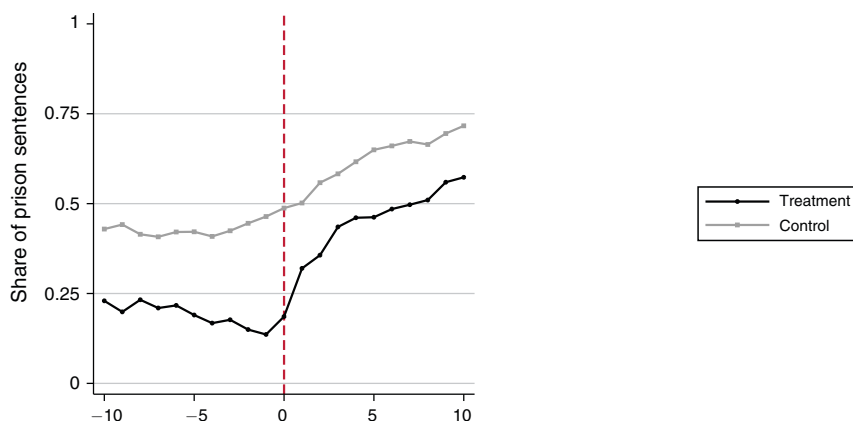


FIGURE 4. SENTENCING AND THE ABOLITION OF CAPITAL PUNISHMENT

*Notes:* The figure shows the annual share of convicted cases in the treatment (black) and control (grey) group that were sentenced to death (panel A), transportation (panel B), or prison (panel C) in the ten years before and after the assigned treatment year. The treatment group contains those offenses for which capital punishment is abolished and the control group those that are never or always capital eligible (see Table 1 for details). The vertical line marks the offense-specific year of abolition of capital punishment for offenses in the treatment group. The reform period for the control group is a weighted average by the share of reforms in each treatment year (see Section IIIA for details).

*Source:* *The Old Bailey Proceedings Online* and own calculations

after the crime-specific year of reform as represented by the vertical line for the treatment group. For the control group offenses, we present a weighted average, where the weights are based on the share of reforms occurring in a given year. Thus, in the year of reform ( $t = 0$ ), the same weight is given to each year in the control group as is implicit in the treatment group; these weights are shifted accordingly in the years leading up to and following the reforms. The share of death sentences in panel A is fairly steady in the treatment group (around 35 percent) leading up to its abolition. In the first complete post-reform year, the share sentenced to death drops to zero. In the control group, the death sentence share is just over 0 percent both before and after the reform; it is not zero due to the “always capital” offenses. Panel B demonstrates the

substitution from capital punishment to transportation for only the treated offenses in the year immediately after the reform and parallel patterns for treatment and control offenses prior to (and after) the reform. Specifically, transportation levels for the treatment group jump from around 40 percent (lower than the controls) to around 60 percent (substantially higher than the controls) in the reform year; thereafter, there is a downward trend in transportation for both treatment and control offenses as society becomes disillusioned with transportation. Finally, as seen in panel C of Figure 4, incarceration in both the treatment and control groups is fairly constant in the years leading up to the reforms, and more prevalent (about 20 percentage points) in the control group. Incarceration jumps in the reform year for the treatment group, partially closing the gap in its use as a sanction. Thereafter, incarceration rates are increasing in a parallel fashion for both groups, consistent with anecdotal evidence on the timing of the rise of imprisonment as a preferred sanction.<sup>25</sup>

### B. Empirical Methodology—Capital Punishment

Motivated by the variation across offenses in the timing of the abolition as well as a number of offenses for which capital punishment status does not change, we adopt a difference-in-difference design to estimate the effect of the reforms on the chance of conviction. We estimate the baseline specification, presented in equation (1), for the years 1803 to 1871:

$$(1) \quad GV_{iot} = \alpha + \beta_1 noncapital_{ot} + \alpha_o + \alpha_t + \alpha_m + \mathbf{X}_{iot} \delta + \epsilon_{iot}.$$

Since the jury has three main verdict choices (guilt of initial charge, guilt of lesser charge, or acquit), we define the dependent variable in two ways. The first considers whether the defendant is guilty of any (initial or lesser) charge; here, jury leniency works solely through acquittals. The second explicitly considers whether the defendant is guilty of the initial charge. Here, the jury has two ways to be lenient: acquit or convict of a lesser offense.<sup>26</sup> Thus,  $GV$  denotes whether defendant  $i$  charged with offense  $o$  in year  $t$  is found “guilty” by the jury. Finally, we consider whether the jury recommends mercy, conditional on conviction; such recommendations can only be observed conditional on a guilty verdict in the first place.

The primary variable of interest, *noncapital*, is an indicator equal to one for offense-year combinations for which the offense is not capital eligible. That is, the treatment indicator *noncapital* turns on upon the abolition of capital punishment for treatment group offenses; for control group offenses, *noncapital* does not change over time and equals zero (one) for always (never) capital offenses. The offense-specific treatment years are reported in Table 1.<sup>27</sup>

<sup>25</sup> Regression results demonstrating a significant positive effect of abolishing capital punishment on the chance of transportation and prison (and no significant differences in the leading years) are available upon request.

<sup>26</sup> We construct a broad variable “guilty of lesser charge” using the detailed verdict variable values: guilty of a lesser offense, manslaughter (different from the genuine offense category manslaughter), or guilty for a theft under a certain value below the value originally charged.

<sup>27</sup> Capital punishment for sodomy and wounding was abolished in stages; our baseline uses the first change as the reform year.



The baseline difference-in-difference specification includes: offense fixed effects ( $\alpha_o$ ) to control for baseline differences in case characteristics and conviction rates across offenses; year fixed effects ( $\alpha_t$ ) to capture other criminal justice reforms that affected all offenses; month fixed effects ( $\alpha_m$ ) to capture seasonality in criminal behavior and even jury behavior (given the absence of heat, it is certainly feasible that deliberations were different in the summer and winter); and a vector of controls ( $\mathbf{X}$ ) including the defendant's gender, the number of defendants, and the defendant's criminal history in subsample analyses.

Standard errors are clustered on the specific offense level. Although this results in a low number of clusters (25 clusters overall and fewer in crime type subsamples), we choose that specification to allow for within offense correlation in the error term, which is important in the presence of serial correlation (Bertrand, Duflo, and Mullainathan 2004). However, following Angrist and Pischke (2009), we demonstrate the robustness of our results to a number of alternative clustering strategies: one-way and two-way cluster for offense by year as well as offense and session, as well as block bootstrapping and a wild  $t$ -bootstrapping procedure (Cameron, Gelbach, and Miller 2008) to correct for the small number of clusters when clustering by offense.<sup>28</sup>

For  $\beta_1$  to represent the causal effect of the abolition of capital punishment on conviction rates, we make the usual parallel trends assumption, i.e., that the change in conviction rates for treatment offenses would have been the same as that for control offenses in the absence of the reforms. Figure 5 presents the share of jury decisions resulting in conviction for any offense (panel A) and the original charge (panel B) for the treatment and control offenses (the latter is again weighted by the share of reforms in a given year) and is suggestive of parallel pre-reform trends. Pre-reform conviction rates for both groups are fairly flat in the years leading to the reforms. We will more formally test for parallel trends by including offense group and detailed offense-specific time trends and estimating an event-study specification (see Section IIID). Figure 5 also provides the first suggestive evidence that abolishing capital punishment increased conviction rates. In interpreting these figures and comparing them to the following results, it is important to keep in mind that the overall sample disproportionately consists of property offenses.

A causal interpretation relies on four additional assumptions. First, abolishing capital punishment must have changed the jury's expectations regarding punishment. Given the active use of pardons, for which we unfortunately do not have data, one can question whether the jury's expectations regarding the chance of capital punishment and an execution actually changed. It is true that the share of death sentences resulting in execution is lower in the early 1800s than previously; King (2006) states that the share sentenced to death who were actually hanged in London dropped from 48 percent from 1784–1790 to 10 percent in 1816–1818. While a 10 percent execution rate is not in itself insignificant, it is not clear that jurors of the time would even be aware of such a drop given that there were actually “as many people hanged in London in the 1820s as in the 1790s” due to a higher number of

<sup>28</sup> See, e.g., Cameron and Miller (2015) or Mackinnon and Webb (2017) for a discussion of inference when the number of clusters is small, which can lead to problems even in bootstrapping procedures.

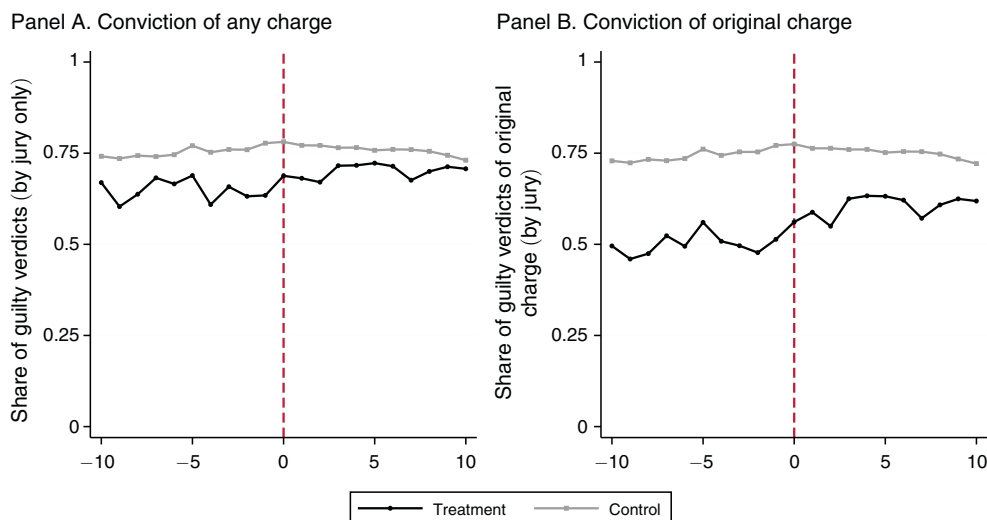


FIGURE 5. CONVICTION RATES AND THE ABOLITION OF CAPITAL PUNISHMENT

*Notes:* The figure shows the annual share of guilty verdicts (cases convicted by jury) in the treatment (black) and control (grey) group for all offenses in the sample in the ten years before and after the assigned treatment year and relative to all cases tried by jury. The treatment group contains those offenses for which capital punishment is abolished and the control group those that are never or always capital eligible (see Table 1 for details). Panel A defines conviction as guilty of any offense while panel B defines conviction as guilty of the original charge. The vertical line marks the offense-specific year of abolition of capital punishment for offenses in the treatment group. The reform period for the control group is a weighted average by the share of reforms in each treatment year (see Section IIIA for details).

*Source:* *The Old Bailey Proceedings Online* and own calculations

capital indictments (King 2006, p. 102). In fact, Gatrell (1994) writes that though the annual hanging rate of convicted capital offenders was cut in half from 1805 to 1818, the average annual number of executions rose from 63 to 103. Moreover, the share of convicted capital cases sentenced to death, as recorded in the proceedings and read by the public (including future jurors), remains fairly constant between 30 percent and 40 percent in the years leading up to the reforms (see Figure 4). That said, it cannot be ruled out based on anecdotal evidence that the jury had already lowered their expectations on the likelihood of a judge sentencing a capital defendant to death or increased the expectation on the chance of a defendant receiving a pardon. If this were the case, then we would likely underestimate the effect of abolishing capital punishment on jury verdicts (since the jury would have internalized at least some of this expectation in its decisions prior to the reform). The fairly flat conviction rates leading up to the reforms, however, and the lack of an effect in the “leading” years would suggest that this is not a significant empirical concern.<sup>29</sup>

<sup>29</sup> More generally, what is the effect of repealing a punishment that is hardly ever used? Can this affect jury behavior? It would seem fair to say that this would depend on the severity of the punishment in question: capital punishment is so severe that even the abolition when the chance of such punishment is very low can have a large effect on the expected punishment.

Second, although the abolition of capital punishment was doubtfully a “random” policy, our strategy assumes that the *timing* of the offense-specific abolition was random. It took more than 40 years to abolish capital punishment for all treated offenses. And, the only crime-specific movement that we are aware of is forgery, for which English bankers pushed to have the death penalty abolished (Hans and Vidmar 1986). The absence of a change in conviction (Figure 5) or sentencing (Figure 4) behavior in the years immediately preceding the reforms supports the assumption that jurors and defendants did not know what offense would be reformed next. We formally demonstrate the lack of anticipatory effects in our event-study analysis. However, to the extent that anecdotal evidence suggests a lead effect for forgery, we demonstrate the robustness of the overall results in online Appendix Figure 3 to excluding forgery.

Third, we assume that the composition of cases and associated quality of evidence presented to the jury did not change after the reform; otherwise, an increase in conviction rates due to the reforms would be observationally equivalent to one due to improved evidence. We discuss potential channels, e.g., criminal activity or defendant pleas, through which this assumption may be violated in Section IIIE and empirically test whether the quality of evidence changed.

Finally, a causal interpretation relies on our ability to disentangle the effects of abolishing capital punishment from anything else happening at the time. We believe there are a number of reasons to justify this claim. First, other criminal justice reforms—the introduction of the police in 1829, the shifting burden of proof to the prosecution in 1827, or the entitlement to defense attorneys in 1836—applied to all offenses at the Old Bailey. Simply including year effects controls for any reforms that affect all offense categories. To the extent one does not believe this to be the case—although we have no anecdotal evidence to suggest this—much of our analyses are conducted at the broad offense category level (e.g., property or violent/sex offenses). It is reasonable to argue that all sub-offenses within these broad categories are equally treated by these reforms. Finally, we note that the abolition of capital punishment did not actually occur for any offense in the same years (1827, 1829, 1836) as the previously mentioned reforms.

### *C. Capital Punishment: Main Results and Robustness Checks*

Table 3 presents the results of estimating equation (1) for all offenses (with and without controls) and the following broad offense categories: property, violent and sex, and fraud. Panel A presents the results for whether the jury convicts of any charge while panel B defines the outcome as conviction of the original charge (thereby grouping a lesser conviction with acquittal). Including our full set of controls (column 2) in panel A, we find that the abolition of capital punishment significantly increases the chance of conviction of any charge by 7.6 percentage points (10.6 percent relative to the mean).

The remaining columns of Table 3 demonstrate the heterogeneity of this effect across crime categories. Abolishing capital punishment increased the chance of conviction of any charge by 22 percentage points (37.0 percent) for violent and sex offenses and 35 percentage points (47.5 percent) for fraud. In contrast, the effect

TABLE 3—BASELINE RESULTS—ABOLITION OF CAPITAL PUNISHMENT AND JURY DECISIONS

Offense:	All (1)	All (2)	Property (3)	Violent and sex (4)	Fraud (5)
<i>Panel A. Guilty of any offense (i.e., original or lesser charge) by jury verdict (0/1)</i>					
Noncapital (0/1)	0.0917 (0.0401)	0.0764 (0.0365)	0.0153 (0.0168)	0.220 (0.0614)	0.345 (0.0648)
Mean	0.720	0.721	0.737	0.595	0.726
Observations	104,910	104,670	83,990	10,017	9,375
Number of clusters	25	25	10	8	3
R <sup>2</sup>	0.053	0.067	0.051	0.107	0.138
<i>Panel B. Guilty of original charge by jury verdict (0/1)</i>					
Noncapital (0/1)	0.171 (0.0342)	0.159 (0.0323)	0.148 (0.0407)	0.158 (0.0764)	0.330 (0.0377)
Mean	0.671	0.671	0.698	0.429	0.702
Observations	104,910	104,670	83,990	10,017	9,375
Number of clusters	25	25	10	8	3
R <sup>2</sup>	0.099	0.109	0.077	0.093	0.158
<i>Panel C. Guilty of lesser offense conditional on guilty by jury verdict (0/1), broad definition</i>					
Noncapital (0/1)	−0.153 (0.0462)	−0.153 (0.0456)	−0.203 (0.0469)	0.0214 (0.131)	0.0017 (0.0272)
Mean	0.069	0.069	0.053	0.280	0.032
Observations	75,571	75,422	61,919	5,961	6,806
Number of clusters	25	25	10	8	3
R <sup>2</sup>	0.256	0.258	0.239	0.221	0.140
<i>Panel D. Recommended for mercy conditional on guilty by jury verdict (0/1)</i>					
Noncapital (0/1)	−0.0590 (0.0194)	−0.0602 (0.0199)	−0.0363 (0.0195)	−0.235 (0.0515)	−0.150 (0.0438)
Mean	0.112	0.112	0.117	0.103	0.076
Observations	75,571	75,422	61,919	5,961	6,806
Number of clusters	25	25	10	8	3
R <sup>2</sup>	0.062	0.066	0.065	0.142	0.054
Offense fixed effect	Yes	Yes	Yes	Yes	Yes
Year fixed effect	Yes	Yes	Yes	Yes	Yes
Month fixed effect	Yes	Yes	Yes	Yes	Yes
Control variables	No	Yes	Yes	Yes	Yes

Notes: The table shows the results for the baseline difference-in-difference regressions corresponding to estimating equation (1) for the sample years 1803–1871. The dependent variable is a dummy variable indicating guilt of any offense (panel A), guilt of the original charge (panel B), guilt of a lesser offense (panel C), and a recommendation for mercy (panel D). Panels C and D are conditioned on the sample of convicted individuals. When included, the controls contain defendant gender and number of defendants. See Table 1 for details on treatment and control group definitions. Standard errors clustered by offense are shown in parentheses below the estimated coefficient.

Source: *The Old Bailey Proceedings Online* and own calculations

for the largest category of property crimes is much smaller (1.5 percentage points or 2 percent) and not statistically significant. The explanation is seen in panel B: abolishing capital punishment increases the overall chance of conviction of the original charge by almost 16 percentage points. For property crimes, juries are almost 15 percentage points (21 percent) more likely to convict of the original charge while the parallel estimates for violent and sex offenses and fraud are 37 percent and 47 percent, respectively. Panel C of Table 3 digs into this further by conditioning on the convicted sample and defining the dependent variable to be conviction of a lesser offense. Abolishing capital punishment reduces the chance of a conviction

of a lesser charge for property offenses but not for violent/sex offenses or fraud. Taken together, panels A–C suggest that the effect of abolishing capital punishment on conviction rates for violent/sex and fraud offenses is driven by a change in the chance of convicting at all; for property offenses, the reform only affects the chance of convicting of the original charge. Finally, panel D presents the impact of abolishing capital punishment on recommendations to mercy for the convicted sample. Overall, column 2 shows that abolishing capital punishment decreased the likelihood of a recommendation to mercy by 6 percentage points (54 percent), which is primarily driven by violent/sex offenses and fraud.<sup>30</sup>

The overall signs and variation in magnitudes of the effects across offense groups are broadly in line with economic intuition. Before abolishing capital punishment, the jury had to find a means of lessening the sentence through the verdict; as capital punishment is abolished, they no longer have to do this. For property crimes, the jury typically acted by convicting of a lesser charge; for instance, the jury could convict an individual of theft of less than 5 shillings to make the offense noncapital (for the point in history when 5 shillings was the threshold). There are fewer violent and sex offenses with corresponding “lesser” offenses (one exception is murder and manslaughter), and thus the jury had to acquit to completely avoid the death sentence. The “recommendation to mercy” results, however, are driven by violent and sex offenses as well as fraud—presumably for those individuals they could not justifiably acquit.

To give some additional perspective on the magnitudes, we first consider the effect of a baseline control—defendant gender. Female defendants are almost 7 percentage points less likely to be convicted than male defendants; thus, the overall effect of the reform (7.6 percentage points) is of a similar magnitude as another variable that is key to explaining jury verdicts. In addition, similarly large or larger effects are seen in modern-day studies of, for instance, prosecutor discretion and jury demographics. For instance, Bjerk (2005) finds that, when constrained by a three-strikes law, prosecutors are almost twice as likely to reduce a felony arrest charge to a misdemeanor. Anwar, Bayer, and Hjalmarsson (2012) show that adding a black to an all-white jury pool can decrease convictions for blacks by 16 percentage points. Adding females to the jury pool in the Central Criminal Court of London (i.e., the same courts studied here) has been found to increase the conviction rate for sex offenses by 17 percentage points (Anwar, Bayer, and Hjalmarsson 2016). Taken together, this suggests that the effect of abolishing capital punishment is large and economically meaningful, but within reason, especially given the magnitude of the reform.

Table 4 presents robustness analyses for conviction of any offense (panels A and D) and the original offense (panels C and F) when looking at all offense categories. Panels B and E presents the same analyses for conviction of any (original or lesser) charge for violent and sex offenses, which as seen in Table 3 is the main offense group driving the result for this outcome. The baseline results from Table

<sup>30</sup> While panels C and D of Table 3 help identify the underlying channels through which juries avoided death sentences (pre-reform) across offenses, they are based on conditional (guilty) samples. Estimating panel C for the unconditional sample finds qualitatively equivalent results (with attenuated but significant estimates). Nonlinear specifications yield qualitatively equivalent results, but suffer from the usual fixed effects concerns.

TABLE 4—ROBUSTNESS ANALYSES—ABOLITION OF CAPITAL PUNISHMENT

Offense:	All				Property	All
	Baseline	Off. group × year FE	Off. group × annual trend	Off. × annual trend	Exclude larceny	Exclude sodomy and wounding
Specification:	(1)	(2)	(3)	(4)	(5)	(6)
<i>Panel A. Guilty of any offense by jury verdict (0/1)</i>						
Noncapital (0/1)	0.0764 (0.0365)	0.0344 (0.0207)	0.0552 (0.0271)	0.0740 (0.0375)	0.0431 (0.0231)	0.0621 (0.0330)
Observations	104,670	104,670	104,670	104,670	31,855	101,909
Cluster	25	25	25	25	9	23
<i>Panel B. Guilty of any offense by jury verdict (0/1), violent/sex offenses only</i>						
Noncapital (0/1)	0.220 (0.0614)	0.205 (0.0649)	0.221 (0.0624)	0.300 (0.0519)	NA NA	0.167 (0.0412)
Observations	10,017	10,017	10,017	10,017		7,256
Cluster	8	8	8	8		6
<i>Panel C. Guilty of original offense by jury verdict (0/1)</i>						
Noncapital (0/1)	0.159 (0.0323)	0.151 (0.0362)	0.154 (0.0327)	0.0955 (0.0271)	0.135 (0.0329)	0.163 (0.0338)
Observations	104,670	104,670	104,670	104,670	31,855	101,909
Cluster	25	25	25	25	9	23
Offense:	All					
	Before 1850	After 1820, before 1850	No multiple defendants	Exclude never death controls	Exclude always death controls	Exclude 3 offenses without law
Specification:	(7)	(8)	(9)	(10)	(11)	(12)
<i>Panel D. Guilty of any offense by jury verdict (0/1)</i>						
Noncapital (0/1)	0.0607 (0.0351)	0.0554 (0.0342)	0.0864 (0.0447)	0.0717 (0.0466)	0.0783 (0.0367)	0.0874 (0.0444)
Observations	86,637	66,679	71,061	42,745	103,989	97,042
Cluster	25	25	25	17	24	22
<i>Panel E. Guilty of any offense by jury verdict (0/1), violent/sex offenses only</i>						
Noncapital (0/1)	0.290 (0.0647)	0.291 (0.0515)	0.218 (0.0822)	0.175 (0.0561)	0.270 (0.0533)	NA NA
Observations	5,625	4,397	6,667	8,034	9,336	
Cluster	8	8	8	5	7	
<i>Panel F. Guilty of original offense by jury verdict (0/1)</i>						
Noncapital (0/1)	0.138 (0.0343)	0.102 (0.0310)	0.194 (0.0290)	0.124 (0.0398)	0.161 (0.0323)	0.184 (0.0292)
Observations	86,637	66,679	71,061	42,745	103,989	97,042
Cluster	25	25	25	17	24	22

Notes: The table shows the results for the robustness analysis corresponding to estimating equation (1) for the sample years 1803–1871 (unless noted otherwise) and as specified at the top of each column. The dependent variable is a dummy variable indicating guilt of any offense (panels A, B, D, and E) and guilt of the original charge (panels C and F) by the jury. All regressions include offense, year, and month fixed effects as well as all control variables. See Table 3 for further details on the baseline specification. Standard errors are clustered by offense and are shown in parentheses below the estimated coefficient.

Source: *The Old Bailey Proceedings Online* and own calculations



3 are shown in column 1. Columns 2 and 3 of Table 4 control for an offense group by year fixed effects and an offense group-specific trend. Column 4 includes a detailed offense-specific time trend. These demanding specifications do not change the qualitative nature of the results.<sup>31</sup> This is in particular true for violent and sex offenses in panel B, for which we find the strongest baseline results. Though the effect size decreases somewhat for the conviction of the original charge outcome (panel C) with offense-specific trends (this could be because a “lesser offense” is not an option for all offense types), the estimate remains quite large and highly significant. Column 5 of Table 4 presents the property crime results when excluding larceny, which (if anything) increases our estimates; see online Appendix Figure 3 for the overall results when leaving out larceny. Larceny is the largest crime category, but also the “messiest”: it was redefined a number of times during our sample and furthermore is not a perfectly clean “control” offense, as the laws did—theoretically—allow for death sentences at various points in history for thefts over a certain threshold. Column 6 demonstrates robustness to excluding sodomy and wounding, which are the two offense categories for which capital punishment was abolished in stages.<sup>32</sup> Restricting the sample to pre-1850 (i.e., before the abolition of transportation) and to post-1820 in columns 7 and 8 yields the same general pattern of results. The remaining columns of Table 4 demonstrate robustness to excluding: multiple defendant cases, never and always capital controls, and simultaneously the three offenses for which we could not find the actual law abolishing the death penalty. Online Appendix Table 3 tests the robustness of our results to alternative specifications of the standard errors and suggests that the baseline results are indeed robust to clustering at a lower level, block bootstrapping, two-way clustering, and a wild *t*-bootstrap.<sup>33</sup>

#### D. Event-Study Test of Parallel Trends and Random Timing Assumptions

This section tests two of the three key identifying assumptions—namely that there are parallel pre-reform trends in conviction rates for treatment and control offenses and that there is random timing of the reforms—in an event-study specification.<sup>34</sup> That is, we interact our treatment indicator with year dummies leading up to and following the reform, going from ten or more years before to ten or more years after the reform year in single-year steps.<sup>35</sup> The omitted year is the year before the first fully treated year, i.e., the year of the reform itself. The results for property and violent/sex offenses are shown in Figure 6, as well as those for all offenses (pooled) in online Appendix Figure 4.

<sup>31</sup> We also estimated column 4 with a quadratic instead of a linear trend. Our conclusions do not change.

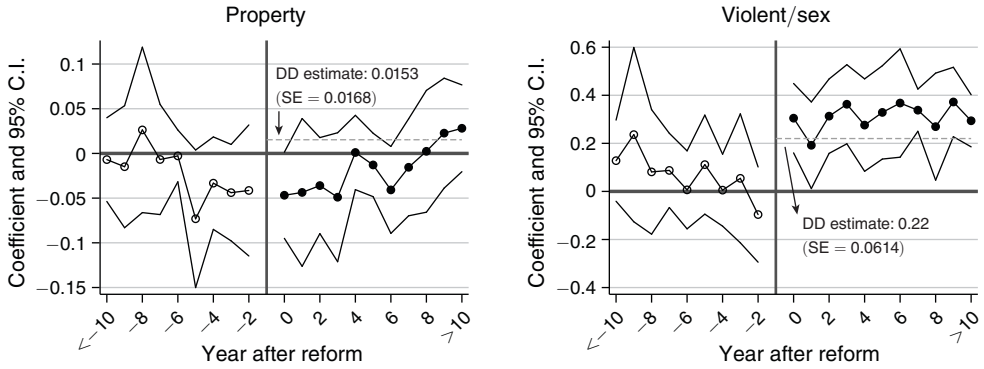
<sup>32</sup> In fact, online Appendix Figure 3 shows that the results are not being driven by any single category.

<sup>33</sup> The table shows the alternative standard errors when pooling all offenses, with and without controls. Note that for the wild *t*-bootstrap, *p*-values instead of standard errors are reported (see Cameron and Miller 2015). To ease comparison, *p*-values for the baseline specification are reported in column 3. Results by the offense group—violent and sex, property, and fraud—are available on request.

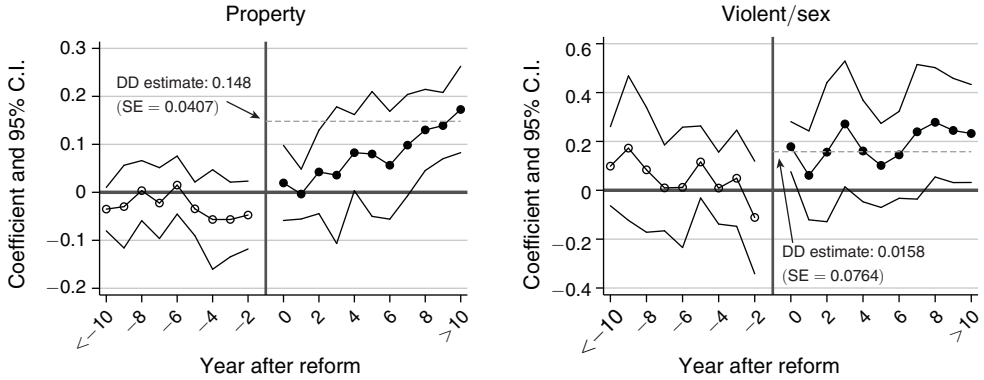
<sup>34</sup> Alternatively, we estimated our baseline with controls for five leads and found the baseline coefficient to be robust to that specification. These results are available in an earlier working paper version (Bindler and Hjalmarsson 2017c).

<sup>35</sup> We chose that window as we restrict our sample to ten years before the first and ten years after the last reform.

Panel A. Conviction of any charge



Panel B. Conviction of original charge



Panel C. Conviction of lesser charge

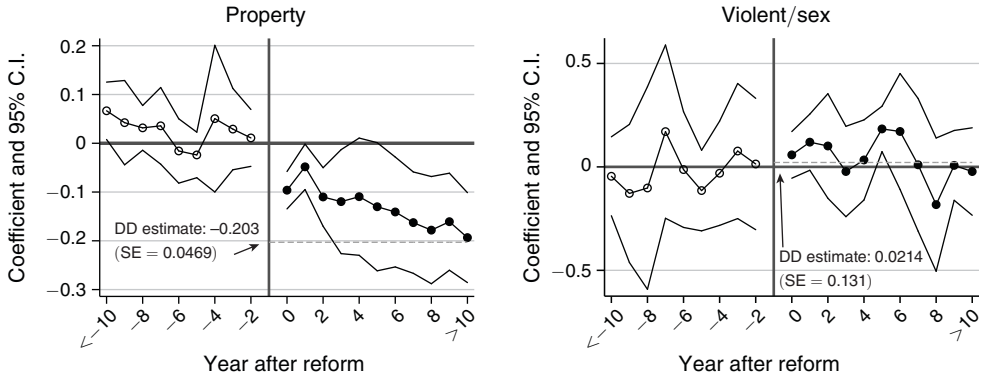


FIGURE 6. EVENT STUDY OF THE ABOLITION OF CAPITAL PUNISHMENT

*Notes:* The figure shows the coefficients and 95 percent confidence intervals from an event-study specification including leads and lags as specified in Section IIID. Panel A shows the results for convictions of any charge, panel B for convictions of the original charge, and panel C for convictions of a lesser charge, each for property and violent/sex offenses. The solid vertical line marks the reference year (year of the reform itself). The dashed horizontal line marks the difference-in-difference estimates as shown in columns 3 and 4 of Table 3. Standard errors are clustered by offense.

*Source:* *The Old Bailey Proceedings Online* and own calculations

Panel A of Figure 6 illustrates the results for convictions of any charge, panel B for convictions of the original charge, and panel C for convictions of a lesser charge. We do not find any differential conviction behavior for treated versus control offenses in the years leading up to the reforms for none of the three outcomes, i.e., the coefficients on the leads in the event-study specification are never significantly different from zero. This supports both assumptions regarding parallel pre-reform trends and no anticipatory effects.<sup>36</sup> The robustness of the results to including offense group-specific trends (and the highly demanding offense-specific trends), as seen previously in Table 4, additionally supports the parallel trends assumption.

The event-study results can also be seen as a robustness check of our baseline difference-in-difference estimates and informative about the timing and persistence of the effects. For violent and sex offenses, there is an immediate and persistent effect on convictions of any charge (panel A). While there is no comparable effect for property offenses for convictions of any charge, we again find a different pattern for both convictions of the original charge (panel B) and convictions of a lesser charge (panel C): there is an immediate effect that increases in size over time. Taken together, these results are in line with the findings in Table 3. For violent offenses, abolishing capital punishment changes the chance of conviction at all, while for property offenses the reforms affect the chance of a conviction of the original or a lesser offense. Further, these effects are very persistent over time. Such permanent effects imply that the baseline effect is not driven by a transitory shock to juries in the reform year (e.g., “upset” about the reforms that could be seen as in favor of the defendants), but indeed captures a persistent change to juries’ behavior.

#### E. Tests of No Confounding Effects Assumption

This section tests the third identifying assumption that there are no confounding effects, i.e., it is only the expected punishment for the charged offense that changes the jury’s behavior and not any other aspects of cases, including the quality of evidence. Specifically, an alternative explanation for our results is that abolishing capital punishment led to an *increase* in the quality of evidence, making it easier for a jury to convict the defendant. Clearly, abolishing the death penalty may impact other agents in the justice system, including potential criminals, defendants, police, and attorneys. But, does it do so in a way that changes the composition of cases facing the jury and, more importantly, in a way that increases the quality of evidence leading to an observationally equivalent increase in conviction rates?<sup>37</sup> We discuss each channel in turn and empirically test for a change in the quality of evidence.

<sup>36</sup> Note that the only exception is fraud, where there is a large reduction in the chance of conviction of the original charge in the year prior to the reform (results not shown). This suggests that juries were acquitting defendants or convicting of lesser charges in the year before the reform, consistent with anecdotal evidence of a push for reform for forgery as discussed earlier. The same exception is found with a difference-in-difference specification that includes leads; it is presented and discussed in the earlier working paper version (see Bindler and Hjalmarsson 2017c).

<sup>37</sup> Indeed, the inherent “selection” that occurs in the multiple stages of the modern judicial process—arrest, charge, bail, trial, and sentencing decisions—has been documented in empirical studies today. That is, early stage decisions have implications for later stage outcomes. With respect to races for instance, Rehavi and Starr (2014) demonstrates that much of the racial gap in sentencing can be explained by racial gaps in earlier stage charge

The economic model of crime (Becker 1968) predicts an increase in the number of crimes upon abolishing the death penalty. Most empirical research testing the deterrent effect of capital punishment today, however, does not suggest this is the case (Donohue and Wolfers 2005). We provide three pieces of evidence that the abolition of capital punishment did not significantly increase the number of Old Bailey cases observed. First, panel A of Figure 7 presents the number of “treated” and “control” cases seen in the ten years before and after the reforms, where the control “reform year” ( $t = 0$ ) is again determined by weighting each year according to the share of reforms in that year. We do not observe a jump in the number of treated cases at the time of the reform, and in fact, the number of control and treated cases appear quite parallel before and after the reform. Second, online Appendix Figure 5 allows the reader to visually inspect the annual number of cases for each treated offense independently, relative to the reform year. There is little to no suggestion of any discontinuous jumps in the number of cases at the reform year, and it is only for rape and wounding where one could potentially argue for a jump over and above the preexisting upward trend (described later) seen for most offenses. Excluding these offenses, however, does not affect the results (see online Appendix Figure 3). What is also apparent from online Appendix Figure 1 (especially for property crimes) is that there is a general trend in the number of Old Bailey cases trialed in the mid-1800s: it is increasing through the 1830s as the catchment area expands and then decreasing through the 1850s as cases shifted to additional (lesser) courts. This nonlinear trend is important to take into account when formally testing whether the abolition of capital punishment increases the number of cases in online Appendix Table 4. Once a quadratic trend is included, there is no effect of the abolition on the number of cases (it is not even marginally significant).<sup>38</sup>

Although the analyses mentioned previously suggest that abolishing capital punishment did not increase the number of cases, it is important to note that this is likely an imperfect proxy for the number of crimes. As in many studies of crime, measurements of crime using official records are a function of the number of crimes reported or caught. Thus, our measure of crime captures both a change in victim reporting as well as criminal behavior. Our results suggest no significant increase in either behavior for most offenses. Further, for offenses such as rape, it could very well be that the small jump observed is due to reporting behavior, since victims may be more likely to report the rapists (often known or related) when capital punishment is off the table.

Another caveat is that even though there is not much action in the number of crimes, the composition of offenders may still have changed. Offenders may be better or worse criminals (e.g., committing sloppier offenses) or of differing socioeconomic status. Such compositional changes could result from both offender behavior and, for instance, police behavior. To the extent we can, we return to this question by looking at the resulting quality of evidence.

---

decisions. With respect to bail and pretrial detention, Dobbie, Goldin, and Yang (2018) demonstrates that such detention increases the chance of conviction.

<sup>38</sup> Note that this is unlikely to be driven by an overly demanding specification: as a benchmark, when we estimate column 4 in Table 4 with a quadratic trend, our baseline results, in particular for violent/sex offenses, hold.

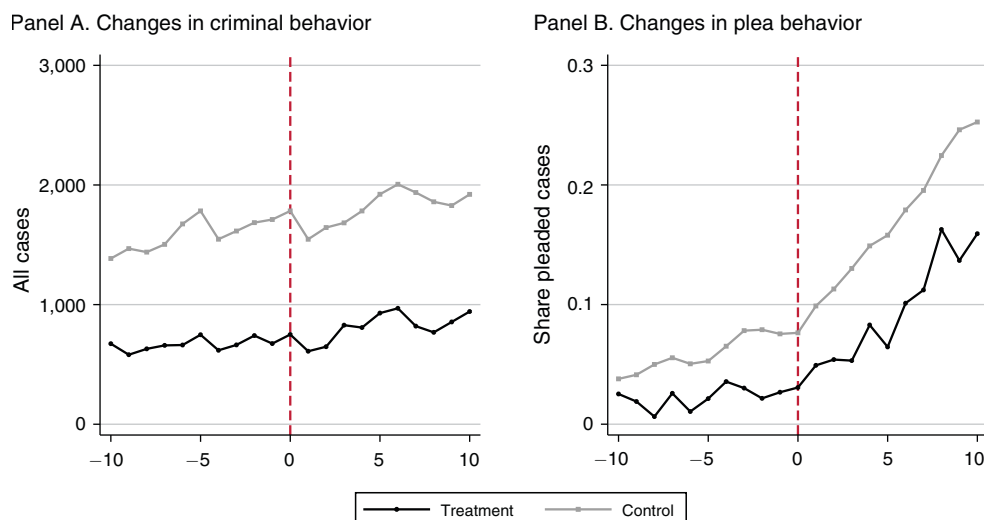


FIGURE 7. CHANGES IN BEHAVIOR AND THE ABOLITION OF CAPITAL PUNISHMENT

Notes: Panel A shows the annual number of cases in the sample tried at the Old Bailey in the treatment (black) and control (gray) group in the ten years before and after the assigned treatment year. Panel B shows the share of pleas in the treatment (black) and control (gray) group. The treatment group contains those offenses for which capital punishment is abolished and the control group those that are never or always capital eligible (see Table 1 for details). The vertical line marks the offense-specific year of abolition of capital punishment for offenses in the treatment group. The reform period for the control group is a weighted average by the share of reforms in each treatment year (see Section IIIA for details).

Source: *The Old Bailey Proceedings Online* and own calculations

Similarly, the change in punishment severity may affect policing behavior or the prosecutors'/grand jury's decision to send a case to trial. For this to be a concern, it must also be that they were bringing cases forward with a differential standard of evidence. Yet, the stakes (punishment severity) decrease with the abolition of capital punishment; thus, one may expect more cases with a *lower* rather than a higher quality of evidence to be brought to trial. This would downward bias our estimates, leading to a *lower* bound of the true effect, but it would not undermine the validity of our results. Finally, though modern-day empirical studies demonstrate that prosecutors respond to harsher punishment, it is not clear that this is a significant concern in this context, since punishments became more lenient and lawyers were a less formalized institution, with the victim primarily serving as prosecutor. It is thus not surprising that Tonry's (1992) anecdotal review of the means by which capital punishment was avoided prior to its abolition highlights the jury and judge pardon decisions, but not lawyers.

Finally, changes in expected punishment may affect a defendant's decision to plead guilty. If this affects the type of cases faced by the jury, then it would raise similar concerns; yet, the most likely scenario is that defendants with the greatest chance of losing (i.e., the strongest evidence against them) are more likely to plead, which would again lower the average quality of evidence of the remaining cases faced by the jury. However, for the most part, pleas did not yet play a large role in the justice system during this period: until 1836, just 3 percent of all cases are recorded

as pleas; after 1836 (contemporaneous with the introduction of defense attorneys for felonies), pleading became more common. Panel B of Figure 7 demonstrates that the share of pled cases trends up in the years surrounding the abolition of capital punishment for both treated and control offenses. One offense group, however, for which plea behavior may be a more substantive concern is fraud: even before 1830, the share of pled fraud cases was greater than 20 percent. One possible explanation is a practice of prosecutors of charging defendants with a first count that is capital (like forgery) and a second that is noncapital (like uttering forged notes). Alschuler (1979) suggests that, in these cases, the defendant pled to the lesser charge and no evidence was presented (by the bank) of the first.<sup>39</sup>

Thus, given the potential importance of changing plea behavior, we check that our results are not driven by a change in case composition due to a change in plea behavior. We reestimate our baseline specifications when including all cases in which a defendant pled guilty and treating them as a conviction (of a lesser offense) by the jury. Accordingly, pleas are treated as convictions when the dependent variable is a conviction of any charge (panel A of Table 5), and essentially treated the same as acquittals when the dependent variable is conviction of the original charge (panel B).<sup>40</sup> Including pleas does not change the qualitative nature of the story: we still see an increase in the chance of any conviction for violent/sex offenses and an increase in the chance of conviction of the original charge for property offenses. Though the magnitudes fluctuate a bit for property and fraud, they remain large and significant, while the violent and sex offense effect does not change at all when including plea cases (due to pleas being much less common for these types of offenses compared to fraud). Indeed, given that some of the largest effects of abolishing capital punishment are seen for offense categories with few pleas, it seems hard to argue that changing plea behavior is driving our results.

We also directly test whether there is a significant change in plea behavior with the reform. That is, panel C of Table 5 reestimates equation (1) when the dependent variable indicates if the defendant pled guilty and indicates that, if anything, there is a decrease in the likelihood of pleading overall and for property offenses. There is a small but only marginally significant increase in the chance of pleas for violent offenses, but this has already been shown to not impact the results. As seen here and in panel B of Figure 7, plea behavior is simply becoming more common at this time for all, and slightly more so for control offenses.

Finally, to more directly assess the bottom line concern—namely that there might be an increase in the quality of evidence with the abolition of capital punishment—we use the Old Bailey online search function to create proxies for the quality of evidence. Specifically, we conduct keyword searches for *evidence*, *police*, and *witness* by year and offense category and normalize by the number of charges in that year (i.e., we look at the hit rate).<sup>41</sup> Panels A–C of Table 6 present the results of

<sup>39</sup> We thank a previous anonymous referee for this reference and highlighting this potential concern.

<sup>40</sup> We only include pleas as a robustness check rather than the baseline since the goal of the paper is to identify the behavior of the jury, and one would have to make assumptions about how to treat a plea (guilty of original charge jury verdict, guilty of a lesser charge jury verdict, or an acquittal).

<sup>41</sup> Specifically, the three searches include the following terms: *evidence*; *witness(es)*; and *policeman*, *police*, *constable*, *watchman*, *watch-man*, *watchmen*, *runner*, *thief taker*, *bobby*, *bobbies*, *peeler*, and *peelers*.



TABLE 5—ROBUSTNESS TO AND TESTS FOR CHANGES IN PLEA BEHAVIOR

Offense:	All (1)	Property (2)	Violent and sex (3)	Fraud (4)
<i>Panel A. Guilty of any offense by jury verdict; plea = guilty (0/1)</i>				
Noncapital (0/1)	0.0471 (0.0282)	0.0047 (0.0150)	0.221 (0.0593)	0.261 (0.0821)
Mean	0.759	0.771	0.612	0.799
Observations	121,410	96,527	10,470	12,808
Cluster	25	10	8	3
R <sup>2</sup>	0.064	0.053	0.111	0.078
<i>Panel B. Guilty of original offense by jury verdict; plea = lesser charge (0/1)</i>				
Noncapital (0/1)	0.248 (0.0387)	0.218 (0.0309)	0.142 (0.0764)	0.362 (0.0135)
Mean	0.578	0.607	0.410	0.514
Observations	121,410	96,527	10,470	12,808
Cluster	25	10	8	3
R <sup>2</sup>	0.092	0.081	0.080	0.128
<i>Panel C. Plea (0/1)</i>				
Noncapital (0/1)	−0.113 (0.0333)	−0.0747 (0.0154)	0.0221 (0.0113)	−0.107 (0.0688)
Mean	0.138	0.130	0.043	0.268
Observations	121,410	96,527	10,470	12,808
Cluster	25	10	8	3
R <sup>2</sup>	0.151	0.174	0.057	0.073
Offense fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Month fixed effects	Yes	Yes	Yes	Yes
Control variables	Yes	Yes	Yes	Yes

Notes: The table shows the results for the robustness regressions corresponding to estimating equation (1) for the sample years 1803–1871 when the dependent variable is a dummy variable in panel A that indicates a guilty jury verdict or a guilty plea for any charge (i.e., treating guilty pleas as equivalent to a guilty jury verdict of any charge) or a dummy variable in panel B that indicates guilt of the original charge, where a guilty plea is treated as jury guilt of a lesser charge. The dependent variable in panel C is whether the defendant pled guilty. See Table 3 for further details on the baseline specification. Standard errors clustered by offense are shown in parentheses below the estimated coefficient.

Source: *The Old Bailey Proceedings Online* and own calculations

estimating equation (1) for each keyword, respectively, for a panel of offense category (26 offenses) by year data, including offense and year fixed effects; panel D uses the aggregated hit rate based on all keywords. When looking at all offenses (column 1), there is a significant reduction in the hit rate for the keywords “evidence” and “witness” after the abolition of capital punishment and a marginally significant reduction in hits on “police.” Similar patterns—though less significant—are seen when looking at violent and sex offenses or property offenses. If anything, we find a decrease (and certainly no increase) in our proxy for the quality of evidence, consistent with the previously mentioned idea that prosecutors, or the grand juries who send the cases to trial, may bring cases with a lower threshold of evidence.

However, it must be noted that there are clear limitations to this approach. First, we cannot distinguish whether the “evidence” is in favor of the defendant or the victim/prosecutor. Second, the keywords themselves may have been used in a

TABLE 6—IDENTIFICATION TEST: ABOLITION OF CAPITAL PUNISHMENT AND QUALITY OF EVIDENCE

Dependent variable: Offenses:	Key word hit rate		
	All (1)	Violent and sex (2)	Property (3)
<i>Panel A. "Evidence"</i>			
Noncapital (0/1)	−0.104 (0.0182)	−0.186 (0.0253)	−0.0369 (0.032)
Observations	1,444	438	557
R <sup>2</sup>	0.386	0.621	0.303
<i>Panel B. "Police"</i>			
Noncapital (0/1)	−0.0375 (0.0218)	−0.0548 (0.0344)	0.0137 (0.0342)
Observations	1,444	438	557
R <sup>2</sup>	0.569	0.823	0.42
<i>Panel C. "Witness"</i>			
Noncapital (0/1)	−0.0688 (0.0217)	−0.059 (0.0363)	−0.0436 (0.0444)
Observations	1,444	438	557
R <sup>2</sup>	0.37	0.597	0.406
<i>Panel D. Aggregated total</i>			
Noncapital (0/1)	−0.210 (0.0458)	−0.299 (0.0737)	0.0668 (0.0822)
Observations	1,444	438	557
R <sup>2</sup>	0.493	0.792	0.361
Offense fixed effects	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes

*Notes:* The table shows the results for the identification test of estimating equation (1) for the sample years 1803–1871 for data aggregated to the offense by year level. The dependent variable is the hit rate corresponding to the key words evidence (panel A), police (panel B), and witness (panel C) and the aggregate total of these three (panel D)—see Section III E for further details on the construction of the variable. See Table 3 for further details on the baseline specification. Robust standard errors are shown in parentheses below the estimated coefficient.

*Source:* *The Old Bailey Proceedings Online* and own calculations

“negated” manner. For instance, consider the word “evidence.” There is clearly a very different implication of a hit that says, “I enter my statement as evidence,” and a hit that says, “There is no evidence.” While we unfortunately cannot say much about the former concern, we can provide some suggestive statistics that the keywords are, in fact, typically used in a positive fashion, as opposed to indicating the “lack” of evidence. Specifically, we use the Old Bailey Corpus website, which is designed to conduct textual analyses of all *spoken* words published in the proceedings from 1810 to 1850, to examine how the word “evidence” is used.<sup>42</sup> We classify the use of the word “evidence” as being affirmative, negative, or unclear. Note that affirmative does not mean the evidence supports guilt, but simply that it is used to indicate

<sup>42</sup> We examined the 150 characters to the left and right of all spoken uses of the word “evidence” on September 5, 2017 in the Old Bailey Corpus website: <http://www1.uni-giessen.de/oldbaileycorpus/search.html>. One must log in to use this search engine.

the existence of evidence. Tabulations of this classification are presented overall by speaker role (defendant, judge, lawyer, victim, and witness) and by decade in online Appendix Table 5. Overall, “evidence” is used in an affirmative manner 80 percent of the time and negative 16 percent of the time. This affirmative use does not differ substantially by decade or speaker role: 74 percent for victims, 79 percent for defendants, 80 percent for witnesses (who comprise the vast majority of uses), 84 percent for judges, and 87 percent for lawyers.

#### F. Capital Punishment: Heterogeneity

Finally, we test for heterogeneity across case characteristics. Table 7 assesses whether the abolition of capital punishment has heterogeneous effects on the chance of conviction of any offense (panel A) or the original offense (panel B) across defendant gender and criminal history. Columns 1 to 3 consider whether the abolition of capital punishment had differential effects for male versus female defendants overall (column 1) as well as for property and violent offenses (columns 2 and 3). Sex offenses are excluded given the lack of female sex offenders. Overall, the abolition of capital punishment increases the chance of any conviction by more than 7 percentage points, with no differential effects by gender. For violent crimes, however, the abolition of capital punishment increases the chance of conviction by 30 percentage points for females and just 18 percentage points (0.305–0.119) for males. That is, a jury’s decision to convict is more sensitive to the abolition of capital punishment for females than males.<sup>43</sup> Panel B indicates that for violent crimes abolishing capital punishment increases the chance of conviction of the original charge for females more than males. Note that while more imprecisely estimated, the point estimates overall and for property crime show a similar pattern. Taken together, this suggests that, prior to the reform, females were more likely to be treated favorably by the jury—they were less likely to be convicted at all and of the original charge for violent offenses.

Criminal history is only recorded after 1832; column 4 demonstrates that the abolition of capital punishment increases the chance of conviction by 10 percentage points for this restricted sample. Controlling for criminal history in column 5 has minimal impact on this finding, even though the variable itself significantly increases the chance of conviction by 28 percentage points. Finally, column 6 suggests that juries had less of a problem imposing a death sentence prior to the reform on individuals of known “bad character,” since the increase in conviction rates caused by abolishing the death penalty is only observed for those without a criminal record. However, the coefficient on the interaction is somewhat imprecisely estimated.

<sup>43</sup> This is consistent with both contemporary and historical evidence that females are treated leniently by the criminal justice system. Bindler and Hjalmarsson (2017d) demonstrates that a gender gap persists in convictions and sentencing at the Old Bailey from 1715 to 1913 while Bodenhorn (2009) finds that females received shorter sentences in nineteenth century Pennsylvania. See, for instance, Starr (2015) for evidence of the gender gap today.

TABLE 7—HETEROGENEITY ANALYSES—ABOLITION OF CAPITAL PUNISHMENT

Offense: Specification:	All	Property Gender	Violent	All	All Criminal history	All
	Interaction (1)	Interaction (2)	Interaction (3)	Baseline after 1832 (4)	Control variable (5)	Interaction (6)
<i>Panel A. Guilty of any offense by jury verdict (0/1)</i>						
Noncapital (0/1)	0.0750 (0.0438)	0.0177 (0.0310)	0.305 (0.0745)	0.100 (0.0457)	0.116 (0.0510)	0.119 (0.0507)
Male defendant (0/1)	0.0663 (0.0171)	0.0773 (0.0163)	0.124 (0.0410)			
Noncapital × male defendant	0.0017 (0.0227)	−0.0031 (0.0221)	−0.119 (0.0322)			
Criminal history (0/1)					0.277 (0.0177)	0.388 (0.0922)
Noncapital × crim. history						−0.112 (0.0920)
Mean	0.721	0.737	0.609	0.727	0.724	0.724
Observations	104,670	83,990	8,702	59,544	57,134	57,134
Cluster	25	10	5	25	25	25
R <sup>2</sup>	0.067	0.051	0.111	0.069	0.105	0.105
<i>Panel B. Guilty of original offense by jury verdict (0/1)</i>						
Noncapital (0/1)	0.190 (0.0421)	0.187 (0.0534)	0.242 (0.0781)	0.118 (0.0369)	0.125 (0.0354)	0.126 (0.0335)
Male defendant (0/1)	0.113 (0.0205)	0.128 (0.0250)	0.141 (0.0169)			
Noncapital × male defendant	−0.0396 (0.0256)	−0.0486 (0.0298)	−0.118 (0.0120)			
Criminal history (0/1)					0.278 (0.0192)	0.294 (0.147)
Noncapital × crim. history						−0.0164 (0.147)
Mean	0.671	0.698	0.437	0.688	0.684	0.684
Observations	104,670	83,990	8,702	59,544	57,134	57,134
Cluster	25	10	5	25	25	25
R <sup>2</sup>	0.109	0.078	0.103	0.102	0.135	0.135
Offense fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Control variables	Yes	Yes	Yes	Yes	Yes	Yes

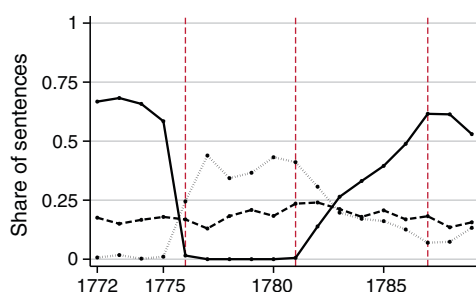
Notes: The table shows the results for the heterogeneity analysis (by criminal history and gender) corresponding to estimating equation (1) for the sample years 1803–1871. The dependent variable is a dummy variable indicating a guilty jury verdict of any offense (panel A) and a verdict guilty of the original offense (panel B). See Table 3 for further details on the baseline specification. Standard errors clustered by offense are shown in parentheses below the estimated coefficient.

Source: *The Old Bailey Proceedings Online* and own calculations

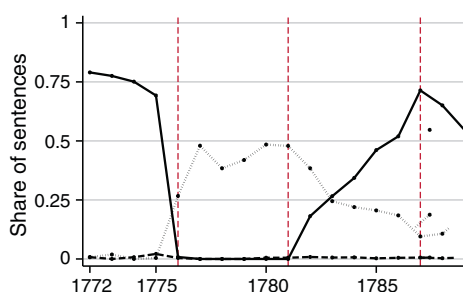
IV. The American Revolution and Temporary Halt of Transportation

This section assesses the impact of the temporary halt of transportation during and following the American Revolution on conviction rates. Figure 8 graphically assesses how punishment changed during the war. The vertical lines correspond

Panel A. All transportation-eligible offenses



Panel B. Noncapital transportation-eligible offenses



Panel C. Capital transportation-eligible offenses

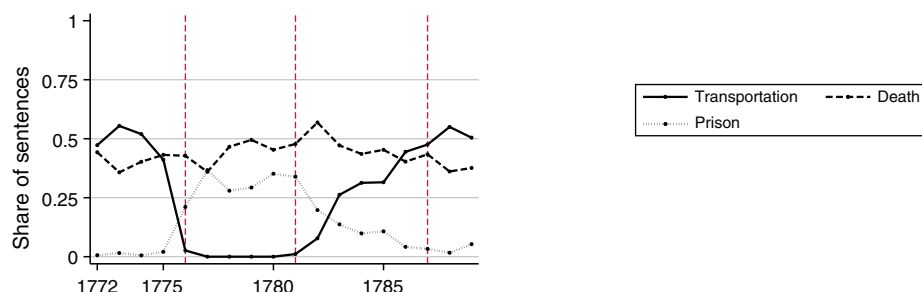


FIGURE 8. SENTENCING AND THE AMERICAN REVOLUTION

Notes: Panel A shows the annual share of all convicted cases in the treatment group (i.e., offenses that were transportation eligible) that were sentenced to transportation (solid line), death (dashed), or prison (dotted) between 1772 and 1789. Panel B shows the equivalent numbers for noncapital-eligible offenses, panel C for capital-eligible offenses. See Table 1 for the transportation-eligible offenses; capital eligibility in 1776 can also be inferred from Table 1. The vertical lines mark the halt of transportation in 1776, the reinstatement in transportation by name only in 1781, and the actual start of transportation to Australia in 1787.

Source: *The Old Bailey Proceedings Online* and own calculations

to 1776 when transportation was first suspended, 1781 when judges began issuing transportation sentences despite the lack of a penal colony, and 1787 when a new Australian penal colony was officially established. Conditioning on guilty cases, panel A presents the share of sentences to transportation, death, and prison for the transportation-eligible offenses (i.e., treated offenses in Table 1). Almost 75 percent of sentences in each year leading up to the war were sentenced to transportation and 0 percent in the years 1776 to 1781. The share of sentences to transportation began to increase again in 1781, until the prewar levels were nearly reached in 1787. Panel A also demonstrates that imprisonment was the primary substitute for transportation, as the share of prison sentences rose from around 0 percent to almost 50 percent during the war while the share of death sentences only rose by 5–10 percentage points. After the war, imprisonment rates decreased again, but not back to zero.

The fact that in some cases transportation was substituted by capital punishment—which is clearly harsher than incarceration—makes it hard to say whether punishment severity (i.e., the jury's expectation regarding punishment) actually increased

or decreased. Yet, it becomes clearer for a subset of offenses upon decomposing the treatment offenses into capital and noncapital eligible. For noncapital offenses (larceny and perjury), the temporary halt of transportation sharply decreased expected punishment with an increase in prison, assuming a prison sentence (despite horrible prison conditions) was perceived as better than transportation to the Americas. For capital offenses, however, both death and imprisonment were substitutes, leaving the change in expected punishment ambiguous (see panels B and C of Figure 8).

A distinguishing feature of this experiment is that the change in punishment severity is driven by a shock—the war—that is exogenous to the justice system. The flip side is that this reduced-form experiment captures not just the sharp first-order effect of the Revolution on transportation but also any other channels through which the war may affect conviction rates.<sup>44</sup> One may be particularly concerned about the immediate aftermath of the war with released military personnel shocking labor markets and tremendous unrest in London following the Gordon Riots in 1780. Unfortunately, we lack a sufficiently large control group to estimate a difference-in-difference specification to isolate the causal effect of the shift in punishment from anything else changing during the war. We are thus limited to using a simple reduced-form pre-post design to provide suggestive evidence regarding the impact of the unexpected (and large) shift in expected punishment in the years during and after the Revolution. Our baseline specification to estimate the effect of the unexpected change in punishment severity upon the temporary halt of transportation in 1776 is presented in equation (2) and focuses on the years 1772 to 1779, i.e., the four years surrounding the start of the war and prior to the riots:

$$(2) \quad GV_{iot}^{cap, noncap} = \alpha + \beta_1 Pre1776_t + \alpha_o + \alpha_m + \alpha_{judge} + \mathbf{X}_{iot}\delta + \epsilon_{iot}.$$

The dependent variable is whether the jury returned a guilty verdict (*GV*) for defendant *i* facing jury *j* charged with offense category *o* in year *t*. The primary variable of interest, *Pre1776*, is a dummy indicating the four years prior to the war—when transportation existed. Defining the specification with the omitted time period having the *changed* expected punishment (no transportation) allows us to expand the same specification to assess the impact of reintroducing transportation. All specifications control for offense and month fixed effects; in this case, more detailed data allow for judge fixed effects. We control for a vector *X* of case-specific characteristics, including defendant gender, number of defendants, and whether the jury (and therefore case) was from London. The latter is important as the Middlesex judges had limited access to the hulks as a potential sentence (Hitchcock and Shoemaker 2015). Prior to 1779, Middlesex had only sentenced 86 prisoners to the hulks, whereas there were at least 600 prisoners in total sentenced to the hulks in the first 20 months after the start of the Revolution. The Penitentiary Act of 1779

<sup>44</sup> See King (2000) for a discussion of the relationship between wars and crime during the 1700s. There is some anecdotal evidence that some convicts were sentenced to serve in the war, but we have no empirical evidence that this occurred to any great extent at the Old Bailey; our data do in fact include military and naval duty as a possible subsentence, and there are only a handful of such sentences (recorded) in the proceedings during the Revolution.



formally eliminated the hulks as an option for the Middlesex justices (Hitchcock and Shoemaker 2015). Anecdotal evidence suggests that a sentence to hard labor on the hulks may have been harsher than a sentence to prison: (i) over one-fourth of prisoners died on the hulks in the first 20 months, (ii) those on the hulks only ate the food provided (“bread and any coarse or inferior food, and water or small beer”) whereas those in prisons were primarily provided food and drink by visiting family and friends, and (iii) those incarcerated on land received frequent visitors. This suggests that the decrease in expected punishment upon the Revolution may have been larger for Middlesex versus London defendants (Hitchcock and Shoemaker 2015).<sup>45</sup>

A number of additional points are worth making about our baseline specification. First, we do not include time trends given that sentencing patterns (share transported/sentenced to death) were relatively constant in the years leading up to the war. Second, we focus on the prewar period. We believe this to be the cleanest natural experiment because in contrast to its reinstatement the halt of transportation was unexpected, our reduced-form framework makes it difficult to disentangle the reintroduction of transportation from general discontent with the justice system (due to overcrowded prisons), and it is difficult to characterize what happened to *expected* punishment in the postwar period when transportation was reinstated in name only. Finally, as denoted in the superscript in equation (2), we divide the transportation-eligible offenses into capital versus noncapital offenses. We emphasize the latter since the halt of transportation unambiguously decreases punishment severity to imprisonment.

The results of estimating equation (2) are presented in Table 8. When considering all transportation-eligible cases and including all controls in column 2, defendants are about 3 percentage points (5.5 percent) less likely to be convicted in the prewar period when transportation is a possible sentence; without controlling for the jurisdiction, we find a slightly negative, but insignificant effect. Columns 3 and 4 decompose these offenses into noncapital and capital, respectively. The overall effect is driven by the noncapital cases, for which punishment severity is unequivocally higher before the halt of transportation; these defendants were almost 5 percentage points (8 percent) less likely to be convicted when transportation was on the table compared to the war period. No effect is seen for capital offenses. Columns 5 and 6 look separately at London and Middlesex cases; the effect of transportation in noncapital cases is larger in Middlesex (about 6 percentage points) compared to London (about 2 percentage points). This is consistent with the larger decrease in expected punishment in Middlesex (due to its limited use of the hulks). Column 7 includes an offense group-specific linear trend; though the coefficient decreases somewhat (and precision is lost), the same qualitative results are seen. Finally, column 8 tests whether the same pattern of results is seen upon the reintroduction of transportation. We expand the sample to 1789 (we stop here as the data are missing from 1790 to

<sup>45</sup> While our data do not distinguish incarceration in prisons from the hulks, we do see that Middlesex defendants were more likely to receive miscellaneous punishments while London defendants received more corporal punishments. Thus, in the face of overcrowded prisons, even the next substitute to transportation appears to be more lenient in Middlesex than London.

TABLE 8—BASELINE RESULTS: AMERICAN REVOLUTION, HALT OF TRANSPORTATION, AND CONVICTIONS

Jurisdiction:	All London and Middlesex cases				London	Middlesex	All cases	
Offense:	All	All	Noncapital	Capital	Noncapital	Noncapital	Noncapital	Noncapital
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Pre-1776 (0/1)	−0.004 (0.014)	−0.031 (0.015)	−0.050 (0.019)	−0.005 (0.023)	−0.023 (0.030)	−0.062 (0.026)	−0.027 (0.037)	−0.048 (0.018)
1780–1786 (0/1)								0.010 (0.022)
Post-1786 (0/1)								−0.025 (0.026)
Mean	0.546	0.564	0.604	0.511	0.703	0.539	0.604	0.631
Observations	5,702	5,420	3,095	2,325	1,227	1,868	3,095	7,794
R <sup>2</sup>	0.062	0.076	0.082	0.067	0.061	0.072	0.084	0.067
Offense fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Judge fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control variables (including jury)	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Off. group-specific linear trends	No	No	No	No	No	No	Yes	No

Notes: The table shows the results for the baseline regressions corresponding to estimating equation (2) for 1772 to 1779 in columns 1–7 and 1772 to 1789 in column 8. See Table 1 for the transportation-eligible offenses; capital eligibility in 1776 can also be inferred from Table 1. The dependent variable is a dummy variable indicating a guilty jury verdict (conviction). Robust standard errors are shown in parentheses below the estimated coefficient.

Source: *The Old Bailey Proceedings Online* and own calculations

1792) and include dummy variables for two additional periods: 1780–1786, capturing the Gordon Riots, its aftermath, and the presence of transportation in name only; and post-1786 when a new Australian penal colony is established. We focus on the latter to avoid confounding our estimates with other channels through which the immediate aftermath of the war may affect crime and conviction rates and because the change in punishment severity is unambiguous—in contrast to 1780–1786. Reinstating transportation (insignificantly) decreases the chance of conviction by about 2.5 percentage points.

In summary, the temporary halt of transportation reduces the chance of conviction for noncapital offenses during the war, when imprisonment in the ship hulks was the primary substitute sentence. While the simple pre-post design limits the causal interpretation, additional findings point in that direction: there is no effect for capital cases, suggesting the results are not driven by a change in attitudes towards all offenders due to the war, and the reverse relationship (though insignificant) is seen upon the reinstatement of transportation.<sup>46</sup>

<sup>46</sup>Transportation was primarily abolished as a sentencing option in the United Kingdom in 1853 (and completely in 1857). At this time, however, expected punishment was already at a low level, with the primary sentencing option being imprisonment. We find no evidence that this change in expected punishment impacted conviction rates (results available upon request). Potential explanations include that either the extent to which changes in punishment severity affect jury behavior depends on the size of the change, or that jurors had already substantially increased conviction rates with the abolition of capital punishment.

## V. Discussion and Conclusion

Using two natural experiments from English history, this paper finds that the decrease in punishment severity resulting from the abolition of the death penalty had a large, significant, and persistent impact on jury behavior, generally leading to the jury being more likely to convict. Similarly, the unexpected decrease in punishment severity at the time of the American Revolution resulted in a significant increase in convictions, albeit one that is smaller than that in the death penalty context. We find that juries differentially applied this behavior depending on the defendant's gender. This (perhaps unintentional) unequal application of justice raises questions about the fairness of the criminal justice system with respect to defendant characteristics, like race and gender, which are clearly still topical today.

Our results provide empirical evidence that juries acted to avoid death sentences before the abolition of capital punishment. This behavior could either be due to a conscious decision to go against clear evidence of guilt due to a disproportionately harsh expected sentence (i.e., jury nullification) or a perhaps less conscious decision to apply a higher threshold when evaluating evidence to determine reasonable doubt when faced with particularly severe punishments. This relates to the discussion of type I (here: false conviction) versus type II (here: false acquittal) errors in judicial decision making. Without knowing a defendant's actual guilt, we cannot empirically disentangle whether this decision is driven by a willingness to allow for type II errors or an aversion to risking type I error, or a combination of both. While the aversion to false convictions appears plausible given the severe punishments at stake (e.g., death), the allowance of false acquittals is in line with anecdotal evidence on jury boycotts when punishments were perceived as too harsh.

Given the historical context, what is the external validity of our analysis to today's criminal justice system? Concerns about type I and type II errors still exist today. False convictions (type I errors) are a common argument brought forward by opponents of the death penalty. One such example is the political and legal debate that preceded the abolition of capital punishment in Illinois, largely motivated by a significant share of wrongful convictions in capital cases (Warden 2012): an editorial of the *Chicago Tribune* stated in March 2007 that "[the] evidence of mistakes, the evidence of arbitrary decisions, the sobering knowledge that government can't provide certainty that the innocent will not be put to death—all that prompts this call for an end to capital punishment."<sup>47</sup> While the Illinois case might be one specific example, numbers by the National Registry of Exonerations support the concern of false convictions, reporting 2,089 exonerations between 1989 and 2016 (819 of which were for homicides) and translating that into more than 18,060 years lost.<sup>48</sup> These numbers illustrate that false convictions are not only a concern with respect to the death penalty, but also with respect to other harsh punishments such as life in prison. The possibility of false acquittals (type II errors) also clearly plays a role in

<sup>47</sup> See [http://articles.chicagotribune.com/2007-03-25/news/0703250314\\_1\\_death-penalty-death-row-murrah-federal-building-bombing](http://articles.chicagotribune.com/2007-03-25/news/0703250314_1_death-penalty-death-row-murrah-federal-building-bombing), accessed on September 14, 2017.

<sup>48</sup> See <http://www.law.umich.edu/special/exoneration/Pages/Exoneration-by-Year-Crime-Type.aspx>, accessed on September 14, 2017.

courts today, with jurors being excluded from capital cases if they are fundamentally opposed to the death penalty and therefore might engage in jury nullification. Further, while estimates about false acquittals are hard to come by, studies such as the one by Spencer (2007) suggest that they are a relevant concern.

However, the simple existence of type I and II errors today does not necessarily imply that a change in expected punishment today would have the same impact on jury verdicts as it does in our historical context. Rather, it would depend on how the size of the expected costs associated with these errors compares across these two contexts. In a simple framework, one can model the expected cost of an error as a function of the probability of such an error and the associated cost of such a decision. On the one hand, it seems reasonable or likely that the chance of errors today, while nonnegative, is lower than historically given the different institutional contexts (e.g., the allowance of appeals today, extremely short trials with limited evidence historically, and a general shift in the burden of proof from the defendant to the prosecution). On the other hand, the loss associated with such an error may be greater today than historically given the increased moral cost of the death sentence in a society where it is much less accepted. Similarly, juries today may face higher moral costs from noncapital sentences, given the increased knowledge about the severe long-term consequences on the defendant and his/her family. Taken together, the potentially lower probability of an error today combined with its potentially higher cost imply that one cannot conclude whether the expected cost to the jury of such an error is greater or smaller today than historically. Thus, this leaves open the possibility that changes in punishment severity today will impact jury behavior; to make further conclusions, research in a contemporary context is clearly needed.

Finally, our findings highlight the divergence between the policymaker and those tasked with implementing those policies. The success of the former depends on the behaviors and reactions of the latter. This is relevant in the criminal justice system, both respect to juries (as seen in this paper) as well as police, prosecutors, and judges. More generally, this may also be relevant in many nonjudicial contexts, in which individuals may avoid taking actions, e.g., reporting cheating students or reporting households to welfare agencies, that result in what they perceive as a disproportionate “punishment.”

## REFERENCES

- Alschuler, Albert W. 1979. “Plea Bargaining and Its History.” *Columbia Law Review* 79 (1): 1–43.
- Angrist, Joshua D., and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist’s Companion*. Princeton: Princeton University Press.
- Annual Probation Survey and Annual Parole Survey. 1980–2014. Bureau of Justice Statistics. <https://www.bjs.gov/index.cfm?ty=dcdetail&iid=271> (accessed June 6, 2016).
- Annual Survey of Jails. 1980–2014. Bureau of Justice Statistics. <https://www.bjs.gov/index.cfm?ty=dcdetail&iid=261> (accessed June 6, 2016).
- Anwar, Shamena, Patrick Bayer, and Randi Hjalmarsson. 2012. “The Impact of Jury Race in Criminal Trials.” *Quarterly Journal of Economics* 127 (2): 1017–55.
- Anwar, Shamena, Patrick Bayer, and Randi Hjalmarsson. 2016. “A Jury of Her Peers: The Impact of the First Female Jurors on Criminal Convictions.” National Bureau of Economic Research (NBER) Working Paper 21960.
- Beattie, J. M. 1986. *Crime and the Courts in England, 1660–1800*. Oxford, UK: Oxford University Press.

- Becker, Gary S.** 1968. "Crime and Punishment: An Economic Approach." *Journal of Political Economy* 76 (2): 169–217.
- Bentley, David.** 1998. *English Criminal Justice in the Nineteenth Century*. London: Hambledon Press.
- Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan.** 2004. "How Much Should We Trust Differences-In-Differences Estimates?" *Quarterly Journal of Economics* 119 (1): 249–75.
- Bindler, Anna, and Randi Hjalmarsson.** 2017a. "Path Dependency in Jury Decision-Making." [https://www.dropbox.com/s/9m228k5eosg1uz6/ab\\_rh\\_path\\_dependency\\_v12\\_submitted.pdf?dl=0](https://www.dropbox.com/s/9m228k5eosg1uz6/ab_rh_path_dependency_v12_submitted.pdf?dl=0).
- Bindler, Anna, and Randi Hjalmarsson.** 2017b. "Prisons, recidivism and the age-crime profile." *Economics Letters* 152: 46–49.
- Bindler, Anna, and Randi Hjalmarsson.** 2017c. "The Fall of Capital Punishment and the Rise of Prisons: How Expected Sentences Affect Jury Verdicts." Centre for Economic Policy Research (CEPR) Discussion Paper 11888.
- Bindler, Anna, and Randi Hjalmarsson.** 2017d. "The Persistence of the Criminal Justice Gender Gap: Evidence from 200 Years of Judicial Decisions." Unpublished.
- Bindler, Anna, and Randi Hjalmarsson.** 2018. "How Punishment Affects Jury Verdicts: Evidence from Two National Experiments: Dataset." *American Economic Journal: Economic Policy*. <https://doi.org/10.1257/pol.20170214>.
- Bjerk, David.** 2005. "Making the Crime Fit the Penalty: The Role of Prosecutorial Discretion under Mandatory Minimum Sentencing." *Journal of Law and Economics* 48 (2): 591–625.
- Bodenhorn, Howard.** 2009. "Criminal sentencing in 19th-century Pennsylvania." *Explorations in Economic History* 46 (3): 287–98.
- Bushway, Shawn D., Emily G. Owens, and Anne Morrison Piehl.** 2012. "Sentencing Guidelines and Judicial Discretion: Quasi-Experimental Evidence from Human Calculation Errors." *Journal of Empirical Legal Studies* 9 (2): 291–319.
- Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller.** 2008. "Bootstrap-Based Improvements for Inference with Clustered Errors." *Review of Economics and Statistics* 90 (3): 414–27.
- Cameron, A. Colin, and Douglas L. Miller.** 2015. "A Practitioner's Guide to Cluster-Robust Inference." *Journal of Human Resources* 50 (2): 317–72.
- Census of Jail Inmates.** 2005. Bureau of Justice Statistics. <https://www.bjs.gov/index.cfm?ty=dcdetail&iid=404> (accessed June 6, 2016).
- Cook, Chris, and Brendan Keith.** 1975. *British Historical Facts 1830–1900*. Palgrave Historical and Political Facts. London: Palgrave Macmillan.
- Devine, Dennis J.** 2012. *Jury Decision Making: The State of the Science*. New York: New York University Press.
- Devine, Dennis J., Kristi M. Olafson, Larita L. Jarvis, Jennifer P. Bott, Laura D. Clayton, and Jami M. T. Wolfe.** 2004. "Explaining Jury Verdicts: Is Leniency Bias for Real?" *Journal of Applied Social Psychology* 34 (10): 2069–98.
- Dobbie, Will, Jacob Goldin, and Crystal S. Yang.** 2018. "The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges." *American Economic Review* 108 (2): 201–40.
- Donohue, John J., and Justin Wolfers.** 2005. "Uses and Abuses of Empirical Evidence in the Death Penalty Debate." *Stanford Law Review* 58 (3): 791–846.
- Emsley, Clive, Tim Hitchcock, and Robert Shoemaker.** 2013. "Punishments at the Old Bailey." *Old Bailey Proceedings Online*, April. [www.oldbaileyonline.org](http://www.oldbaileyonline.org).
- Feeley, Malcolm M.** 1997. "Legal Complexity and the Transformation of the Criminal Process: The Origins of Plea Bargaining." *Israel Law Review* 31 (1–3): 183–222.
- Flower, Shawn M.** 2008. *Disparities in Jury Outcomes: Baltimore City vs. Three Surrounding Jurisdictions—An Empirical Examination*. Abell Foundation. Baltimore, September.
- Freedman, Jonathan L., Kirsten Krismier, Jennifer E. MacDonald, and John A. Cunningham.** 1994. "Severity of Penalty, Seriousness of the Charge and Mock Jurors' Verdicts." *Law and Human Behavior* 18 (2): 189–202.
- Garrett, Brandon, Daniel Krauss, and Nicholas Scurich.** 2017. "Capital Jurors in an Era of Death Penalty Decline." *Yale Law Journal Forum* 126: 417–30.
- Gatrell, V. A. C.** 1994. *The Hanging Tree: Execution and the English People 1770–1868*. Oxford: Oxford University Press.
- Hans, Valerie P., and Neil Vidmar.** 1986. *Judging the Jury*. Cambridge, MA: Perseus Publishing.
- Hitchcock, Tim, and Robert Shoemaker.** 2015. *London Lives: Poverty, Crime and the Making of a Modern City, 1690–1800*. Cambridge, UK: Cambridge University Press.



- Hitchcock, Tim, Robert Shoemaker, Clive Emsley, Sharon Howard, Jamie McLaughlin, et al.** 2013. "The Old Bailey Proceedings Online, 1674-1913, Version 7.1." *Old Bailey Proceedings Online*, April. [www.oldbaileyonline.org](http://www.oldbaileyonline.org).
- Horrell, Sara, Jane Humphries, and Ken Sneath.** 2015. "Consumption conundrums unraveled." *Economic History Review* 68 (3): 830-57.
- Iyengar, Radha.** 2011. "Who's the Fairest in the Land? Analysis of Judge and Jury Death Penalty Decisions." *Journal of Law and Economics* 54 (3): 693-722.
- Kaplan, Kalman J., and Roger I. Simon.** 1972. "Latitude and Severity of Sentencing Options, Race of the Victim, and Decisions of Simulated Jurors: Some Issues Arising from the 'Algiers Motel' Trial." *Law and Society Review* 7 (1): 87-98.
- Kelly, Morgan, and Cormac Ó Gráda.** 2016. "Adam Smith, Watch Prices, and the Industrial Revolution." *Quarterly Journal of Economics* 131 (4): 1727-52.
- King, Peter.** 2000. *Crime, Justice, and Discretion in England 1740-1820*. Oxford: Oxford University Press.
- King, Peter.** 2006. *Crime and Law in England, 1750-1840*. Cambridge, UK: Cambridge University Press.
- Lacasse, Chantale, and A. Abigail Payne.** 1999. "Federal Sentencing Guidelines and Mandatory Minimum Sentences: Do Defendants Bargain in the Shadow of the Judge?" *Journal of Law and Economics* 42 (S1): 245-70.
- Langbein, John H.** 1987. "The English Criminal Trial Jury on the Eve of the French Revolution." In *The Trial Jury in England, France, Germany: 1700-1900*, edited by Antonio Padoa Schioppa, 13-39. Berlin: Duncker and Humblot.
- Lee, Jean N.** 2017. "The Process is the Punishment: Juror Demographics and Administration in State Courts." *American Law and Economics Review* 19 (2): 361-90.
- Lehmann, Jee-Yeon K., and Jeremy Blair Smith.** 2013. "A Multidimensional Examination of Jury Composition, Trial Outcomes, and Attorney Preferences." [http://www.uh.edu/~jlehman2/papers/lehmann\\_smith\\_jurycomposition.pdf](http://www.uh.edu/~jlehman2/papers/lehmann_smith_jurycomposition.pdf).
- Mackinnon, James G., and Matthew D. Webb.** 2017. "Wild Bootstrap Inference for Wildly Different Cluster Sizes." *Journal of Applied Econometrics* 32 (2): 233-54.
- National Prisoner Statistics Program.** 1980-2014. Bureau of Justice Statistics. <https://www.bjs.gov/index.cfm?ty=dcdetail&iid=269> (accessed June 6, 2016).
- Philippe, Arnaud, and Aurélie Ouss.** Forthcoming. "'No Hatred or Malice, Fear or Affection': Media and Sentencing." *Journal of Political Economy*.
- Rehavi, M. Marit, and Sonja B. Starr.** 2014. "Racial Disparity in Federal Criminal Sentences." *Journal of Political Economy* 122 (6): 1320-54.
- Schefflin, Alan, and Jon Van Dyke.** 1980. "Jury Nullification: The Contours of a Controversy." *Law and Contemporary Problems* 43 (4): 51-115.
- Spencer, Bruce D.** 2007. "Estimating the Accuracy of Jury Verdicts." *Journal of Empirical Legal Studies* 4 (2): 305-29.
- Starr, Sonja B.** 2015. "Estimating Gender Disparities in Federal Criminal Cases." *American Law and Economics Review* 17 (1): 127-59.
- Starr, Sonja B., and M. Marit Rehavi.** 2013. "Mandatory Sentencing and Racial Disparity: Assessing the Role of Prosecutors and the Effects of Booker." *Yale Law Journal* 123 (1): 2-80.
- Texas Judicial Branch.** 2014. *Annual Statistical Report for the Texas Judiciary: Fiscal Year 2014*. Austin: State of Texas.
- Tonry, Michael.** 1992. "Mandatory Penalties." *Crime and Justice* 16: 243-73.
- Ulmer, Jeffery T., Megan C. Kurlychek, and John H. Kramer.** 2007. "Prosecutorial Discretion and the Imposition of Mandatory Minimum Sentences." *Journal of Research in Crime and Delinquency* 44 (4): 427-58.
- Vickers, Chris, and Nicolas L. Ziebarth.** 2016. "Economic Development and the Demographics of Criminals in Victorian England." *Journal of Law and Economics* 59 (1): 191-223.
- Vidmar, Neil.** 1972. "Effects of decision alternatives on the verdicts and social perceptions of simulated jurors." *Journal of Personality and Social Psychology* 22 (2): 211-18.
- Voth, Hans-Joachim.** 1998. "Time and Work in Eighteenth-Century London." *Journal of Economic History* 58 (1): 29-58.
- Warden, Rob.** 2012. "How and Why Illinois Abolished the Death Penalty." *Law and Inequality* 30 (2): 245-86.