The civilizing process in London's Old Bailey

Article (Published Version)

The civilizing process in London's Old Bailey

Sara Klingenstein[a], Tim Hitchcock[b], and Simon DeDeo[a,c,1]

[a]Santa Fe Institute, Santa Fe, NM 87501; [b]Department of History, University of Sussex, Brighton BN1 9RH, United Kingdom; and [c]Department of Informatics, School of Informatics and Computing, Indiana University, Bloomington, IN 47408

The jury trial is a critical point where the state and its citizens come together to define the limits of acceptable behavior. Here we present a large-scale quantitative analysis of trial transcripts from the Old Bailey that reveal a major transition in the nature of this defining moment. By coarse-graining the spoken word testimony into synonym sets and dividing the trials based on indictment, we demonstrate the emergence of semantically distinct violent and nonviolent trial genres. We show that although in the late 18th century the semantic content of trials for violent offenses is functionally indistinguishable from that for nonviolent ones, a long-term, secular trend drives the system toward increasingly clear distinctions between violent and nonviolent acts. We separate this process into the shifting patterns that drive it, determine the relative effects of bureaucratic change and broader cultural shifts, and identify the synonym sets most responsible for the eventual genre distinguishability. This work provides a new window onto the cultural and institutional changes that accompany the monopolization of violence by the state, described in qualitative historical analysis as the civilizing process.

cultural evolution | group cognition | social systems | bureaucracy | information theory

Over the course of the 19th century two developments helped to shape the modern world. First, the Western nation–state took on a newly bureaucratic form, including a newly regulated system of law and justice (1–3). And, second, the levels of violence in Western societies plummeted to a unique, all-time low. (For an overview of the statistical evidence for the long-term decline in violence, see ref. 4. Most recently, this literature and argument forms the basis for ref. 5.) Throughout Western Europe, murder became a rarity, whereas the judicial systems of both Europe and North America took on a new professionalism marked by comprehensive record keeping. Scholars using traditional historical methodologies have described this decline in violence as the central component of what is called the "civilizing process." (This is building on ref. 6. For the specifically British experience, see refs. 7–10.)

Part of a formal theory of cultural development designed to explain the emergence of the modern Western state, this civilizing process is taken to include a wide variety of forms of interpersonal relationships ranging from the rise of the concept of politeness to the relationships between classes. The core claim of the theory is that the state effectively monopolized the use of violence over the course of the 16th to 20th centuries, becoming an important actor in both the control of the cultures that encouraged violence, and in the direct policing and control of violence itself.

The bureaucracies that characterize this shift undertook information gathering on an unprecedented scale, designed in part to inform later decision making, and the digitization of these records makes it newly possible to study the civilizing process in a quantitative fashion. The data here come from the detailed records of the Central Criminal Court, or Old Bailey, in London (11, 12). The Old Bailey has heard trials for serious crimes in London and the surrounding counties since the 16th century, and forms one of the longest-running bureaucracies in the modern Western world.

We analyze the 112,485 trial records, encompassing more than 20 million (semantic) words of testimony recorded between 1760 and 1913, a period during which trial reports were at their most comprehensive. We focus on the lexical semantics of spoken testimony: the meaning-laden words used by speakers that can be grouped as synonyms at different thresholds of similarity. Our methods allow us to study the explicitly named semantic structures of these texts over more than 2 orders of magnitude in resolution, from the word-stem level (2.6×10^4 categories) to a synonym set level, with 1,040 categories, to a highly coarse-grained representation with only 116 categories.

We report two major findings. Our first finding is the emergence, by the beginning of the 19th century, of an increasingly clear distinction, within the record of spoken language, between trials associated with violent and nonviolent indictments. The result amounts to the creation of a new bureaucratic genre (13, 14) concerned with the management of violence, and its emergence is followed by a long-term, secular trend over the next 100 y that reinforces the genre through increasing distinctiveness of trial types over time. The increasingly distinctive form of statements associated with the management of violent crime provides strong quantitative evidence for the emergence of new social norms and cultural controls on behavior as context for the decline of violence over the course of the 19th century.

Our methods allow us to identify the particular categories most responsible for the distinction, and to determine their changing influence over time. Based on this analysis, our second finding concerns the large-scale structure of the genre's creation. Rather than the amplification of a particular initial pattern, we find significant change in the nature of the distinctions drawn. The civilizing process amounts not only to a strengthening of prior norms, but also to their elaboration and change over decade- and century-long timescales. Inspection of these most salient categories allows us to associate this process with both bureaucratic innovation associated with the civilizing process and

Significance

One of the characteristics of the modern era is the emergence of new bureaucratic and social mechanisms for the management and control of violence. Our analysis of 150 y of spoken word testimony in the English criminal justice system provides new insight into this critical process. We show how, beginning around the 1800s, trials for violent and nonviolent offenses become increasingly distinct. Driven by a shifting set of underlying signals, this long-term shift in the underlying norms of the system involves both changes in bureaucratic practice and in civil society as a whole.

ANTHROPOLOGY

PSYCHOLOGICAL AND COGNITIVE SCIENCES

changes in the culture at large, and to determine the timing and strength of factors associated with these two distinct drivers.

Methods

For the period between 1674 and 1913, trials heard at the Old Bailey were also published in more or less detail in a publication popularly known as *The Old Bailey Proceedings*. This series has since been digitized, incorporating the classification of all crime types, as the Old Bailey Online (OBO) (11). The first 86 y of the corpus have not been used as they are marked by gaps in the archive, inconsistent and frequently truncated reporting, and periods of inconsistent censorship (15–19). We start our analysis in 1760, which isolates a period during which the publication of transcribed testimony was common practice. The data become even more reliable after 1778, when the City of London required that the transcripts provide a "true, fair and perfect narrative" (ref. 17, p. 468). As well as including words spoken during the trial, as put down by a short-hand reporter, the set also includes extensive metadata, including charge, verdict, and punishment.

A separate layer of mark-up, drawn from "The Old Bailey Corpus" (OBC) (12), has been used to distinguish statements that purport to be transcriptions of spoken testimony, separating out these records from the surrounding text. An example of a trial, that of Mary Hall for Grand Larceny, with verbatim spoken testimony highlighted, is shown in *SI Appendix*. Excepting a short period of censorship focused on acquittals in the early 1790s, and a longer-term practice of censoring graphic details from cases involving sexual offenses, the tagged speech is believed to provide a verbatim record of nearly every word spoken in the courtroom (20).

We map trials to probability distributions over their lexical semantics in a bag-of-words model: In particular, we coarse-grain the words that appear in the trial into named categories defined by similarity of meaning. To do this, we rely on an explicit structure provided by the 1,040 synonym classes of the 1911 edition of *Roget's Thesaurus* (21), originally created over a 47-y period between 1805 and 1852 (22). As an example, a trial in the OBC may include the word "purse," tagged as a noun, in the spoken testimony. "Purse" is found in two synonym sets: set 800, "money" and set 802, "treasury," and the weight of this single appearance is spread over two entries. Words with similar semantics also fall into these categories (the noun "coin" is also found in set 800, for example). Coarse-graining at this level allows us to deal with a more manageable set of 1,040 categories, rather than the 2.6×10^4 possible word stems, at the cost of losing fine-grained distinctions that might exist between stealing someone's purse and stealing someone's coins.

Roget's classification scheme is a nested hierarchy, and we are able to coarse-grain at different levels. On the level just above the synonym sets there are 116 categories. Continuing our example, both sets 800 and 802 coarse-grain into the superordinate category "monetary relations," which also includes the synonym sets associated with "wealth," "poverty," "credit," and "debt" (among others). As before, we are able to measure the statistics of these coarse-grained sets with greater reliability, at the cost of losing fine-grained distinctions between (for example) the theft of a purse and the failure to pay a debt.

The utility of this hierarchy for coarse-graining relies on the fact that Roget's original classifications, and his organization of these classifications into larger groups, mirrors human performance in determining the degree of relatedness between the meanings of words (23–25). In empirical and comparative studies, Roget's matched human-level abilities on classification tasks (26), at a level comparable to other databases such as WordNet (27, 28). In addition to tracking human performance, use of the explicit semantic representation provided by Roget's provides categories fixed independently of corpus frequency. [Unsupervised methods, such as topic modeling via latent Dirichlet allocation (29), can provide a different and complementary window onto this process, including the study of both nonsemantic pattern (30) and unmarked semantic distinctions within indictment classes.]

We focus on the distinction between violent and nonviolent crimes. Using the metadata attached to each trial we divide our corpus into two classes, based on the category of indictment: violent trials (with offense categories such as "assault" or "murder") and nonviolent trials (with offense categories such as "fraud" and "theft"). Our classification for violent crimes is defined in terms of the OBO's assigned offense category and subcategory; offenses classed as "violent" include murder, assault, rape, kidnapping, and wounding, and are listed in full in *SI Appendix, section S6*.

We then compute the distinguishability of trial classes by means of the Jensen–Shannon divergence (JSD) between their feature vectors. The JSD has a particularly simple interpretation as the amount of information gained about the trial class (violent vs. nonviolent) given a single sample from the trial itself. From the Shannon axioms, it has a unique functional form, which can be written as the sum of two Kullback–Leibler (KL) divergences:

$$\text{JSD}\left(\vec{p},\vec{q}\right) = \frac{1}{2}\left[\text{KL}\left(\vec{p}\left|\frac{1}{2}\left(\vec{p}+\vec{q}\right)\right.\right) + \text{KL}\left(\vec{q}\left|\frac{1}{2}\left(\vec{p}+\vec{q}\right)\right.\right)\right], \qquad [1]$$

where $\vec{p}$ and $\vec{q}$ are the two probability distributions, associated with the two distinct classes. As an information-theoretic quantity, JSD behaves sensibly under coarse-graining (31). This means we can ask about the distinguishability of trials given the appearance of a synonym set (the fine-grained level), or given only knowledge of the superordinate category that the synonym set falls under. The latter quantity is always less than or equal to the former; if fine-grained information is particularly useful for distinguishing the trial types, then coarse-graining will reduce the JSD significantly; conversely, if subtle distinctions are unimportant, then the JSD for the coarse-grained categories will be close to the fine-grained case.

To characterize the divergence between the violent and nonviolent trials, we measure each category's individual contribution to the JSD. This allows us to identify the categories that are most responsible for the divergence. The partial KL,

$$\text{KL}_i\left(\vec{p},\vec{q}\right) = p_i \log \frac{2p_i}{(p_i+q_i)} \qquad [2]$$

tells us the extent to which feature i in the data are a reliable signal of the $\vec{p}$ distribution. These information-theoretic methods allow us to mathematize the central question of this paper: to what extent the ways of talking in a criminal trial differed between violent and nonviolent offenses, and how these distinctions changed over time.

Results

Genre Emergence. Fig. 1 shows the long timescale history of violent and nonviolent trial distinguishability in the OBC at the synonym set level and at the more coarse-grained category level. Immediately visible is the steady increase, starting around the end of the 18th century, in the distinguishability of violent vs. nonviolent trials. Our use of the JSD measure allows us to translate this directly into information gained by knowledge of the semantics of a single word—"friend," say, or "money"—heard during the course of a trial.

For comparison, we show the 1σ ranges for a null-class model, where the 60 possible offense categories are split into two null-class genres of size equal to the violent and nonviolent genres. The null allows us to establish that until the last decade of the 18th century, the distinction between violent and nonviolent crimes simply did not exist: It is no easier to distinguish these two trial classes than it is to distinguish an arbitrary split.

Fig. 1 thus establishes not just the development, but the very emergence at these levels of coarse-graining of two distinct genres governing violent and nonviolent offenses. This takes place around roughly 1790 for the fine-grained (synonym set) case and 1795 for the coarse-grained case.

We find that a linear-increase model at both levels of coarse-graining is strongly preferred to deflationary accounts of the Old Bailey that ascribe the increase in distinguishability to purely administrative changes. Focusing our attention on the years after 1778, when transcript data are recorded directly in shorthand and are most reliable, jurisdiction changes occur in 1834 (partial extension to the whole of England) and 1855 (less serious crimes reallocated to Magistrates courts). Although these discrete events affect both the size of the court and the relative numbers of violent and nonviolent indictments, our measures of distinguishability are insensitive to volume. As an additional check for systematics, we test our model against an alternative deflationary model where the court has three stationary epochs. This alternative model is decisively rejected in favor of the linear increase model at $p < 10^{-12}$ (*SI Appendix, section S1* and additional checks in *SI Appendix, section S2*).

Strong evidence for the emergence and development of these distinct trial genres can also be found within the two classes of violent and nonviolent indictments. We break the violent indictment class into "violent theft," "assault," and "killing"; and the nonviolent class we can break into "theft," "deception"
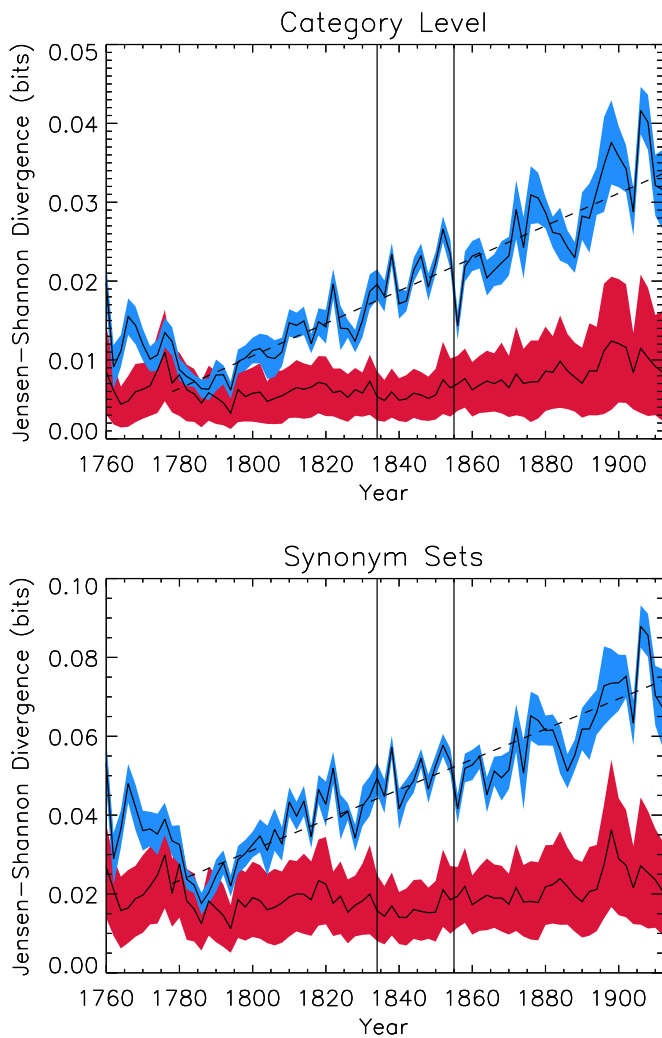
Category Level

Synonym Sets

Fig. 1. The emergence of bureaucratic distinctions between violent and nonviolent offenses in the Old Bailey, 1778–1913. Shown is the JSD (Eq. **1**) at the category level (*Upper*) and the synonym set level (*Lower*). The blue line shows the JSD (and 1σ ranges) for the violent–nonviolent distinction, whereas the red line shows the ranges for a null hypothesis where the trial genres are defined by arbitrary groupings of indictment classes. Overplot as a dashed line is the maximum likelihood fit for a linear increase in the JSD beginning at 1778. Vertical lines indicate the two major administrative changes in the system in this period.

(including fraud, forgery, perjury, and bigamy), and a miscellaneous "other" category.

We can then see how trials for, say, assault and theft become increasingly distinct over time. Table 1 presents these results at the category level coarse-graining. We find significant evidence of a long-term increase in distinctiveness against nonviolent offenses for both violent theft and assault. This increasing distinctiveness across the violent–nonviolent divide contrasts with distinctions internal to the nonviolent–violent genre, which show evidence for the opposite process: As trials for violent and nonviolent offenses become more distinct, trials for (in particular) theft and deception, and theft and "other nonviolent," become less distinct over time (*SI Appendix*, section S5).

The one exception to the increasing distinctiveness of violent and nonviolent offenses occurs in the case of trials for killing. One explanation for this is an account of the civilizing process where the judicial system has already distinguished these most serious offenses—including murder, manslaughter, and infanticide—

Table 1. Increasing distinctiveness between indictment classes

Category (violent)	Category (nonviolent)	Slope, bits per decade	SD*
Violent theft	**Theft**	0.0011	+20.1
	Deception	0.0022	+14.6
	Other	0.0012	+8.5
Assault	**Theft**	6×10^{-4}	+6.2
	Deception	0.0026	+14.2
	Other	6×10^{-4}	+3.6
Killing	Theft	-3×10^{-4}	−3.1
	Deception	0.0013	+6.8
	Other	-1×10^{-4}	−0.8

*Linear model parameters and significance (SDs away from zero) for indictment class pairs at the category level coarse-graining. Pairs with significant evidence of increasing distinctiveness (>5σ) are marked in boldface.

from other forms of crime before the 1780s. We find that the distinction between trials for killing and for theft in 1780 is already larger than that between violent theft and theft in 1910.

Sources and Statistics of Distinguishability. To gain a complementary view into the cultural practices and institutions driving the distinctions established in *Genre Emergence*, we can use the partial KL, Eq. **2**, to examine the changing words and synonym sets that mark a trial as violent or nonviolent. Fig. 2 shows this graphically for the top 24 sets, which amount to a significant minority (~30%) of the total JSD found in the data. As can be read from Fig. 2*A*, for example, the synonym set "death" is a characteristic and distinguishing feature of the violent genre throughout the time period we have under investigation; similarly, "greatness" (best understood as a generic hyperbole) persists until the end of the 1890s.

As an aid to interpretation, Table 2 shows the top five words (by partial KL) for each of the 12 synonym sets whose dynamical evolution is shown in Fig. 2. Sets are listed in order of contribution to the average partial KL. In a few cases (4 of 60)—noted by an asterisk in Table 2—the ambiguity of the word stem is significant ("arms" stems to "arm," for example, and so instances are indistinguishable from reference to the body part).

Fig. 2 shows that the secular rise of distinguishability is not simply the amplification of a particular initial pattern; rather, the actual synonym sets that serve to distinguish violent from nonviolent trials are themselves changing. We can order sets by average year, weighted by partial KL. A set like "death," that remains relevant throughout, has a centroid of 1835, nearly exactly at the center of our date range. Some sets, such as "clothing," "receptacle," and "arms" are characteristic of earlier years, whereas others, such as "remedy" (associated with medical evidence) appear later.

This ability to track long timescale shifts in relevance allows us, among other things, to trace both the changing tolerance of violence and a concern with less lethal methods. The sets "impulse" (including terms such as knock, hit, and strike) and "sharpness" (including references to knives, razors, and blades) come to prominence as signatures of violent trials in the second half of the 19th century.

This particular shift can be decomposed into two underlying components: changes in the overall frequency of use and increasing predictive power independent of frequency. Comparing the first 20 y after 1778 with the last 20 y of our data, both "sharpness" and "impulse" nearly double in frequency, whereas "arms" (which includes references to pistols, revolvers, and bullets) declines slightly, by about 10%. These shifts are associated with the OBC itself, and not changing prevalence in the wider culture. In the British English Google Books corpus, the top five words associated with "sharpness" and "impulse" decline by 5%

(a) Partial KL, Violent to Mixture

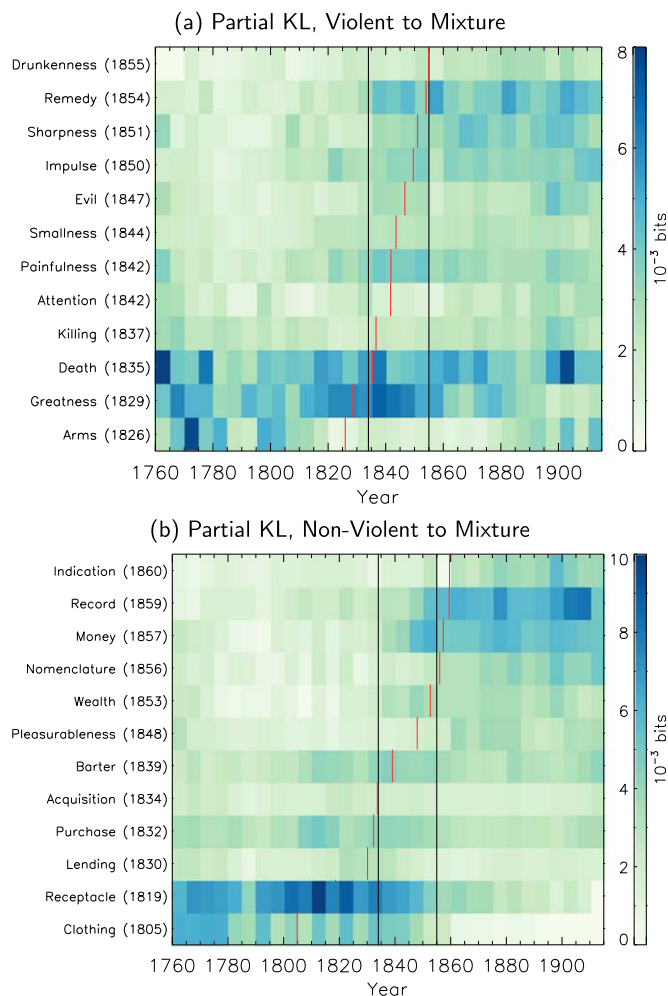

(b) Partial KL, Non-Violent to Mixture

Fig. 2. Distinguishing features of violent and nonviolent trials in the OBC. Shown here is the partial KL, KL_i (Eq. 2), for the 12 most distinctive synonym sets for violent indictments (*A*) and nonviolent indictments (*B*). Sets low on the y axis are associated with earlier dates. The centroid for each term is shown in parentheses and marked in red.

and 13%, respectively, between the two periods, whereas "arms" declines by 32% (32).

Independent of these overall frequency shifts, the increasing focus on less lethal methods as "violence of concern" can be quantified by means of the conditional partial KL, defined as the partial KL for category i, KL_i, divided by the probability of category i, p_i. Conditional partial KL can be interpreted as the information in bits for the trial type, given that the word seen was from category i. In the first 20 y after 1778, "impulse" provides 0.50 bits of information in favor of an indictment for violence and "sharpness" 0.42 bits. In the last 20 y of our data this rises to 0.79 bits and 0.78 bits, corresponding to a drop of roughly a factor of 3.3 and 3.9 in error rate, respectively. "Arms" sees a much smaller increase; it is already more predictive (0.68 bits) in this early period and increases to 0.77 bits, a drop in error rate of only 35%. Information in bits for a binary choice can be characterized equivalently as an error rate, corresponding to the probability of the less likely outcome (in this case, a nonviolent indictment).

Fig. 2B shows the synonym sets that most characterize the nonviolent genre across time. "Clothing" and "receptacle" are most prominent in marking nonviolent trials from 1760 through the mid-1840s. However, beginning in the mid-1840s and continuing

Table 2. Distinguishing features at the synonym set and lexical level for violent and nonviolent indictments

Synonym set	Most significant lexical instances
Violent	
Death	Deceased (adj., n.), death, die, dead
Greatness	Very, so, much, great, most
Remedy	Hospital, surgeon, doctor, sister, medical
Painfulness	Blow, wound, cut (v.), pain, hurt
Impulse	Knock, hit, kick, strike, push
Arms	Pistol, revolver, arms,* bullet, shot
Sharpness	Knife, razor, point, sword, blade
Killing	Blood, stab, kill, murder (n., v.)
Evil	Wound, bruise, injury, mischief, scratch
Smallness	Handful,* inch, little (adv., adj.), mouthful*
Attention	See, examine, observe, notice, attend
Drunkenness	Drunk, drink (v., n.), beer, tipsy
Nonviolent	
Receptacle	Office, bag, pocket, basket, till
Record	Book, note, sign (v.), entry, write
Purchase	Shop, buy, pay, purchase (v., n.)
Money	Money, bill, pocket, cash, coin
Barter	Shop, ask, business, transaction, custom
Clothing	Housing,* things, drawers, coat, shawl
Nomenclature	Name (n., v.), entitle, description, term
Wealth	Money, property, fortune, wealthy, afford
Pleasurableness	Good, desire, transport, satisfactory, satisfy
Indication	Letter, note, sign (v.), bill, ticket
Lending	Pawnbroker, pawn, note, loan, let
Acquisition	Find, receive, produce, take, pay

adj., adjective; adv., adverb; n., noun; v., verb.
*The ambiguity of the word stem is significant.

through the end of our data, synonym sets such as "money," "record," and "indication" (the latter two including references to financial documents) become more characteristic of non-violent trials. These shifts in semantic structure of nonviolent trials go hand in hand with the shifts in the kinds of nonviolent crime considered at the Old Bailey; they are roughly coterminous with the jurisdiction shift in 1855, but the relationship is not clean: strong signals in "record" and "money," for example, accompany an earlier diversification of indictment types (*SI Appendix*, section S4).

Discussion

The 154 y included in our analysis, from 1760 to 1913, span unprecedented political and social change, including the American and French Revolutions, the Napoleonic wars, the Industrial Revolution, Britain's transition around 1850 to majority urbanization, and the rise of her worldwide Empire. Remarkably, despite these transformative events in civil society, the treatment of violence in the legal bureaucracy is a story of gradual, long-term change, rather than one marked by abrupt transitions.

It is a process that cannot be attributed to any one or more of the small number of discrete policy changes that punctuate the history of the court. The causes associated with this phenomenon must act on a timescale much longer than any individual policy development (years), or even the career of an individual bureaucrat or political leader (years to decades). These long-term shifts are visible not only in the emergence of a binary distinction between violence and nonviolence, but also in more fine-grained distinctions such as those between theft and violent theft, or deception and physical assault.

Although we can track the (declining) per capita murder and manslaughter rate in our data, and our results confirm those of prior studies (*SI Appendix*, section S3), direct knowledge of

historical violence—particularly when it does not rise to the level of murder—is hard to come by. One of the most important features of our work is a new ability to track, quantitatively, changing attitudes toward less serious forms of violence.

Having eliminated confounds such as jurisdictional changes, as well as changing language use in society at large and trial length and words per trial, we are able to quantify an essential correlate of the civilizing process: not a concern for the welfare of individuals per se, but rather the creation of new bureaucratic practices and distinctions associated with the control of violence. Changing attitudes become visible to our analysis in both (*i*) how the underlying events of a crime are turned into an indictment, and (*ii*) how those events are presented and discussed at trial.

The sifting of crime into different categories by the criminal justice system (theft vs. assault, for example) was the result of a complex negotiation between the victim, the police (after the formation of the Metropolitan Police in 1829), the clerk of the Grand Jury, the court, and the defendant. The classification of behavior into indictments that underlies the developing genre distinction we find here represents a consensus about the nature of the crime, reached through a complex system of negotiation and bureaucratic decision making. The gradual evolution of a public prosecution service and the emergence of a lawyer-led system both contributed to the emergence of a new adversarial trial form. For example, the Prisoners' Counsel Act in 1836 gave defendants new privileges and their counsel an increasing role in the trial process. The "lawyerization" of the court room and professionalization of the system of prosecution began in the 1690s as the state began to fund the prosecution of the most serious cases, but the introduction of expenses for prosecutors in a series of gradual steps (1752, 1778, and 1826) encouraged the increasing use of professional lawyers hired by the victim. It was only in the middle of the 19th century that the police gradually took on the role of prosecutor (ref. 33 but see also refs. 34–38).

Punishment also underwent significant change as a result of a series of legislative interventions. Before the 1820s, when confronted with a crime involving both theft and violence, prosecuting the theft could result in just as serious a punishment as prosecuting the violence. Indeed, until 1823 even minor theft cases could result in just as serious a punishment as prosecuting violence. However, by the 1840s, prosecuting for the violent features of the crime could result in more serious retribution. This changing legal landscape may also be the source of a relative increase in prosecutions for "wounding." That crimes previously treated as nonviolent are now shifted to the violent category does not, however, end up blurring the distinctions. Even as less serious violence comes under prosecution, the genre distinction becomes stronger, indicating that the civilizing process happens in both the categorization and discussion of violence.

Meanwhile, discussion of the crime was already influenced before trial by the process of preparing depositions and establishing a strategy. Testimony heard by the court was in many respects a rehearsed statement, at least in part influenced by the professions involved in the system. Coaching by lawyers in how to present a crime could easily influence the language used to the jury. And the early decades of the 19th century witnessed a remarkable growth in the role of lawyers and, as importantly, lawyers' tendency to use theatrics as a part of their professional practice (39), tracked in part by the "greatness" (generic hyperbole) synonym set.

Some of our signals are driven solely by the emergence of new topics. In the 1830s, for example, we see the emergence of "remedy" as a significant signal of the violent indictments. This synonym set contains medical terminology and appears around the time doctors became frequent witnesses in trials for both assault and murder, providing forensic testimony that includes medical terms. The practice of calling doctors as expert witnesses is localized primarily in violent trials in which the body forms a major site of evidence (40).

A combination of both categorization and discussion effects appears to explain the changing roles of firearms, blades, and "impulse" (beating), showing the development of a legal bureaucracy that comes, increasingly, to discuss non-firearm violence, and to treat that violence as a matter of direct concern, rather than in passing during a trial for a nonviolent offense.

The markers of nonviolent trials are also changing. As the court came to focus more fully on "serious" felony cases, to the exclusion of minor theft (normally prosecuted in magistrate's courts from 1855), the synonym sets involved underwent a corresponding change. Early trials for nonviolent crimes prominently include the semantics of clothing, words such as "pocket" and "basket," and the semantics of pawnbrokers, all associated with the theft and fencing of small, portable items of value stolen from one's person.

From around the midcentury, however, nonviolent trials are increasingly marked by terms in the "record" category. This includes words such as "signature," "certificate," and "ledger," reflecting the increasing importance of fraud in the mix of Old Bailey trials, and relative changes in the bureaucracy of justice itself (words such as "register" and "record" reflect evidence given by a legal officer referring to notes). In contrast to both the violent–nonviolent distinction and the internal structure of the violent indictments, these shifts internal to the nonviolent genre may be driven by the changing mix of indictment types that appeared before the court.

Conclusions

Our analysis suggests three conclusions. First, that the cultural framing of violence was a gradual process driven by evolving social attitudes and not by any particular legislative or bureaucratic change. As the work of several generations of observers, 593 working in this tradition have suggested (1–3), the civilizing process was a deep-rooted and multivalent phenomenon that accompanied the growing monopoly of violence by the state, and the decreasing acceptability of interpersonal violence as part of normal social relations. Our work here is able to track an essential correlate of this long-term process, most visibly in the separation of assault and violent theft from nonviolent crimes involving theft and deception.

The second conclusion is that a number of distinct processes combined to form the distinct genres that emerged around 1780, translating an inchoate social phenomenon into a measurable cultural process. What underlay this process was not the amplification of an originary distinction. Instead, the grounds for the distinction itself appears to be undergoing significant change. Changing strategies available to victims, decreasing tolerance of milder forms of violence, and the increasing role of words associated with the new bureaucracy itself have each been identified as significant contributors to this central genre distinction.

The third conclusion concerns the use of archives of state to identify not simply the nature and effect of intended policy changes, but also the often unintended and less controllable social and bureaucratic changes that drive the long-term evolution of culture. Our analysis uses an explicit semantic structure to coarse-grain the lexical semantics of texts, and uses information theory to quantify the resulting patterns in terms of bounds on inference. This allows us to give name and number to the *deus ex machina* that animate historical change in a fashion that allows for comparison across long timespans and between traditions.

1. Webb S, Webb B (1906) *English Local Government* (Longmans, Green and Company, London), Vol 1.
2. Joyce P (2013) *The State of Freedom: A Social History of the British State Since 1800* (Cambridge Univ Press, Cambridge, UK).
3. Daunton M (2008) *State and Market in Victorian Britain: War, Welfare and Capitalism* (Boydell Press, Woodbridge, Suffolk, UK).
4. Eisner M (2003) Long-term historical trends in violent crime. *Crime and Justice* 30: 83–142.
5. Pinker S (2011) *The Better Angels of Our Nature: Why Violence Has Declined* (Penguin Group, New York).
6. Elias N (2000) *The Civilizing Process: Sociogenetic and Psychogenetic Investigations*, eds Dunning E, Goudsblom J, Mennell S (Wiley, Hoboken, NJ), 2nd Ed.
7. Gatrell VAC (1980) *Crime and the Law: The Social History of Crime in Western Europe Since 1500*, eds Gatrell VAC, Lenman B, Parker G (Europa, London), pp 238–337.
8. Stone L (1983) Interpersonal violence in English society 1300–1980. *Past and Present* 101(1):22–33.
9. Wiener M (2006) *Men of Blood: Violence, Manliness, and Criminal Justice in Victorian England* (Cambridge Univ Press, Cambridge, UK).
10. Emsley C (2005) *Hard Men: The English and Violence Since 1750* (Hambledon and London, London).
11. Hitchcock T, et al. (2012) The Old Bailey Proceedings Online, 1674–1913. Version 7.0. Available at www.oldbaileyonline.org. Accessed March 24, 2012.
12. Huber M, Nissel M, Maiwald P, Widlitzki B (2012) The Old Bailey Corpus. Available at www.uni-giessen.de/oldbaileycorpus. Accessed June 5, 2013.
13. Guillory J (2004) The memo and modernity. *Crit Inq* 31(1):108–132.
14. Poovey M (2008) *Genres of the Credit Economy: Mediating Value in Eighteenth- and Nineteenth-Century Britain* (Univ of Chicago Press, Chicago).
15. Emsley C, Hitchcock T, Shoemaker R (2012) *Publishing History of the Proceedings*. Available at www.oldbaileyonline.org/static/Publishinghistory.jsp. Accessed June 5, 2014.
16. Devereaux SPR (2003) The fall of the sessions paper: The criminal trial and the popular press in late eighteenth-century London. *Crim Justice Hist* 18(1):57–88.
17. Devereaux SPR (1996) The city and the sessions paper: "Public justice" in London, 1770–1800. *J Br Stud* 35(4):466–503.
18. Harris M (1980) *Sale and Distribution of Books from 1700*, eds Myers R, Harris M (Oxford Polytechnic Press, Oxford).
19. McKenzie A (1998) *Criminal Justice in the Old World and the New: Essays in Honour of J. M. Beattie*, eds Smith GT, May AN, Devereaux S (Centre of Criminology, Univ of Toronto, Toronto).
20. Huber M (2007) Studies in variation, contacts and change English. *Annotating Variation and Change*, eds Meurma-Solin A, Nurmi A (Research Unit for Variation, Contacts and Change in English, University of Helsinki, Helsinki).
21. Roget PM (1991) *Roget's Thesaurus*. Available at www.gutenberg.org/ebooks/22. Accessed June 6, 2014.
22. Hüllen W (2003) *A History of Roget's Thesaurus: Origins, Development, and Design* (Oxford Univ Press, Oxford).
23. Jarmasz M, Szpakowicz S (2004) Roget's thesaurus and semantic similarity. *Recent Advances in Natural Language Processing III: Selected Papers from RANLP 2003, Current Issues in Linguistic Theory*, eds Nicolov N, Bontcheva K, Angelova G, Mitkov R (John Benjamins, Philadelphia), pp 111–120.
24. Aman S, Szpakowicz S (2008) Using Roget's Thesaurus for fine-grained emotion recognition. *International Joint Conference on Natural Language Processing* (Asian Federation of Natural Language Processing, Hyderabad, India), pp 312–318.
25. Kennedy A, Szpakowicz S (2010) Evaluation of a sentence ranker for text summarization based on Roget's Thesaurus. *Text, Speech and Dialogue*, 13th International Conference, eds Sojka P, Horák A, Kopecek I, Pala K (Springer, Brno, Czech Republic), pp 101–108.
26. McHale M (1998) *A Comparison of WordNet and Roget's Taxonomy for Measuring Semantic Similarity*, Proceedings of COLING/ACL Workshop on Usage of WordNet in Natural Language Processing Systems (Association for Computational Linguistics, New Brunswick, NJ), pp 115–120.
27. Kennedy A, Szpakowicz S (2008) *Evaluating Roget's Thesauri*. Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics (The Association for Computer Linguistics, Columbus, OH), pp 416–424.
28. Kennedy A, Szpakowicz S (2012) Fast semantic relatedness: WordNet: Similarity vs. Roget's Thesaurus. *Tiny Transactions on Computer Science*, eds Bailis P, Sherry J (University of California, Berkeley, CA), Vol 1. Available at http://tinytocs.org/vol1/papers/tinytocs-v1-kennedy.pdf. Accessed June 6, 2014 .
29. Blei DM, Ng AY, Jordan MI (2003) Latent Dirichlet allocation. *J Mach Learn Res* 3: 993–1022.
30. Hughes JM, Foti NJ, Krakauer DC, Rockmore DN (2012) Quantitative patterns of stylistic influence in the evolution of literature. *Proc Natl Acad Sci USA* 109(20): 7682–7686.
31. DeDeo S, Hawkins RXD, Klingenstein S, Hitchcock T (2013) Bootstrap methods for the empirical study of decision-making and information flows in social systems. *Entropy* 15(6):2246–2276.
32. Michel JB, et al.; Google Books Team (2011) Quantitative analysis of culture using millions of digitized books. *Science* 331(6014):176–182.
33. Emsley C (2010) *Crime and Society in England, 1750–1900* (Longman, London), 4th Ed, Chap 8.
34. Cairns DJA (1998) *Advocacy and the Making of the Adversarial Criminal Trial, 1800–1865* (Clarendon, Oxford).
35. Gallanis T (1999) The rise of modern evidence law. *Iowa Law Rev* 84(3):499–560.
36. Lemmings D (2000) *Professors of the Law: Barristers and English Legal Culture in the Eighteenth Century* (Oxford Univ Press, Oxford).
37. May AN (2003) *The Bar and the Old Bailey, 1750–1850* (Univ of North Carolina Press, Chapel Hill, NC).
38. Langbein JH (2005) *The Origins of Adversary Criminal Trial* (Oxford Univ Press, Oxford).
39. Devereaux S (2012) Arts of public performance: Barristers and actors in Georgian England. *Crime, Courtrooms and the Public Sphere in Britain, 1700–1850*, ed Lemmings D (Ashgate Publishing, Frahnam, UK).
40. Landsman S (1998) One hundred years of rectitude: Medical witnesses at the Old Bailey, 1717–1817. *Law Hist Rev* 16(3):445–494.

Supplementary Information for "The civilizing process in London's Old Bailey"

Sara Klingenstein [*], Tim Hitchcock [†] and Simon DeDeo [*] [‡] [§]

[*]Santa Fe Institute, Santa Fe, NM 87501, USA,[†]Department of History, University of Sussex, Brighton, BN1 9RH, United Kingdom,[‡]School of Informatics and Computing, Indiana University, 901 E 10th Street, Bloomington, IN 47408, USA, and [§]To whom correspondence should be addressed. Email: sdedeo@indiana.edu

Summary of Supplementary Information

This supplementary information file presents detailed characterization of the Old Bailey over the one-hundred and fifty four years of the data we use.

We first test for two major potential systematic influences: (1) those related to changes in the Old Bailey's jurisdiction, and (2) those related to overall trial volume and number of words.

We then present an account of the ratio between indictments for violent and non-violent crimes and the overall volume of trial accounts, connecting it to both formal changes in Old Bailey jurisdiction and shifts in cultural attitudes.

We then characterize the changing nature of indictments at the Old Bailey, and connect this to both cultural and legal practice.

An additional supplementary analysis is then presented reflecting a result discussed in the main article, *viz.*, that the civilizing process seen at the large-scale level can also be seen at the indictment level. We describe and present detailed graphical and statistical analyses of these sub-processes.

We then present a sample trial from our dataset, with the spoken word testimony highlighted.

Finally, we compare the compare the British English corpus from Google N-grams with the word frequencies found in the Old Bailey Corpus, for the most relevant synonym sets by Partial KL. This allows a year-by-year comparison of word usage as a check against the possibility of systematic and unaccounted-for drift in the use of language at large.

S1: Jurisdiction Tests

Here we examine a crucial "deflationary" explanation for the long-timescale shifts in JSD: that administrative changes in the nature of crimes allocated to the Old Bailey could mimic the genre distinction that this work interprets as being due to large-scale shifts in cultural and bureaucratic norms. If, for example, a trial type is simply shifted to a different court, this could lead to an apparent creation or disappearance of a trial genre without a concomitantly significant change in actual cultural or legal practice. The major administrative shifts that could lead to such an effect in the data are shown as vertical lines in Fig. 1: 1834, when the jurisdiction was extended to the whole of England, and 1855, when less-serious crimes were reallocated to Magistrates courts.[1]

We find that these two dates do not leave a significant, distinguishable signature on the time series, testing this observation via a simple model-selection argument. We consider two models for the post-1778 Jensen-Shannon divergence: the first with two discrete steps at the 1834 and 1855, and a second, linear secular-increase model that does not require dividing the data into epochs. The log-likelihood of the data given the model allows us to reject the administrative-shift model, for the secular-increase model. The log of the maximum-likelihood ratio, $\ln\left[P_{\max}(\text{linear}|\text{data})/P_{\max}(\text{stepwise}|\text{data})\right]$ is equal to +99 for the category level and +102 for the synonym set level. In both cases, the evidence in favor of the secular-

increase model is at a significance, or p value, much less than 10^{-12}.

This result holds even if we use a generalized three-epoch model, where the two dates of transition are allowed to float freely, independently at each coarse-graining level, and without regard to known administrative or historical events. The secular increase model is still strongly preferred ($\Delta\mathcal{L} = +49$ for the category level and $\Delta\mathcal{L} = +29$ at the synonym set level; *i.e.*, still preferred at $p < 10^{-12}$), despite the fact that this generalized model now has five free parameters compared to the two-parameter linear process.

S2: Trial Statistics

We consider only trials with semantically identifiable spoken-word testimony. While most trials contain spoken testimony, cases settled by plea bargain, and hence resulting in a bare record of charge, guilt and punishment, do not. As a result, trials of this sort, which become relatively commonplace after approximately 1820, form the bulk of the data dropped from the OBC, leaving a total of 112,485 trials, and a total of 2.3×10^7 semantically significant words. Figure S3 shows the number of trails available to us once these cuts are made, as well as the average number of words per trial. Not all words are classified in Roget's; determiners such as "the", for example, are without direct lexical semantics. Our analysis works solely with the words assigned semantic significance.

In estimating the relevant information-theoretic quantities of the trials from the data, it is necessary to bias-correct the naively-measured values, and to determine their error ranges; we do this by means of the statistical bootstrap, which preserves the coarse-graining relationship at better than 10^{-4} for the levels of over-sampling we achieve in our data [3]. We do this bootstrap correction by resampling on a trial-by-trial level, a more conservative method that takes into account systematic variation due to within-trial correlations.

The most serious possible contaminant of the strong result we find in Figure 1 is the possibility that over time, as trials became longer and more detailed, more distinctive language was introduced. Fig. S3 allows us to check that the shifts seen in Fig. 1 are not simply due to changes in trial length.

While trial length is positively correlated with Jensen-Shannon divergence (Pearson cross-correlation of 0.52 ± 0.09), the relationship between JSD and trial length is clearly non-monotonic. For example, the Old Bailey of the 1780s devotes more words per trial than in the 1820s, but it is this latter period that shows a strong signal of the genre distinction. It is likely that the same bureaucratic and cultural processes that drive the emergence of the genre distinction are also affected by the time and resources devoted to individual trials, and thus that correlations in the quantities are to be expected.

---

[1] "The Criminal Justice Act", 18 & 19 Vic. c.126 (1855) established summary jurisdiction on a clearly defined basis, allowing people charged with minor theft and other offences to be convicted by two justices. This act was amended only slightly by the Summary Jurisdiction Act, 42 & 43 Vic. c.49 (1879). For a narrative account of the evolution of the bureaucracy of criminal justice in these decades see Ref. [1]. For a statistical approach to the impact of this legislation see Ref. [2].

However, a simple model that assumes a monotonic relationship between trial length and distinctiveness can not explain the results of Fig. 1.

S3: Indictments: Violent vs. Non-Violent

This section considers the change in the number, and ratio, of violent and non-violent cases, and discusses possible influences on our main results.

From 1760 to 1913, the Old Bailey's trials are a snapshot of how the criminal behavior of London was managed by the bureaucracy. Fig. S4 shows the rise, from 1800 to 1840, in the overall number of non-violent indictments (as found in the OBC), and its subsequent decline after 1840. Because trials in the Old Bailey are mostly for non-violent crimes, Fig. S4 has roughly the same shape as the dotted line, reflecting the number of trials overall, in Fig. S3.

Fig. S5 shows the same counts for violent crimes, and Fig. S6 the ratio of violent to non-violent crimes. Between 1760 and 1834, the overall ratio of violent to non-violent crimes declines very slightly—from a high of around 10% in the 1760s to around 5% by 1834 (statistical significance for the negative slope is formally at $+21\sigma$ given bootstrap errors). A major part of this decline results from the fall in the per capita rate of violent crime, which traces the arc of the civilizing process in real-world behavior.

While many violent crimes certainly went unreported, and some jurisdictional and legal changes affected the precise mix of crimes that came to the Old Bailey, the most serious crimes, and in particular, murder, would have always come to the Old Bailey in this period. The murder rate has been recognized as a standard metric for changing levels of violence. While pre-1801 population figures for London are subject to some error, between 1760 and 1815 the population of London nearly doubled, from roughly three quarters of a million to 1.4 million in 1815, rising to three million in 1861, and to over seven million in 1911. The largely unchanging number of murder cases (Fig. S7) heard at the Old Bailey each year translates into a substantial per capita decline. This statistic is complicated by the contemporaneous increase in prosecutions for the lesser crime of manslaughter. This could result from either new perceptions of culpability in accidental death, or a growing leniency on the part of the prosecution, in cases that would previously have been tried as murder. But even in the case of manslaughter, the number of prosecutions per capita of the population was effectively falling. For readers interested in detailed discussion of the overall level of crime in London, see Refs. [4, 5, 6].

In addition to the increasing safety of London, as evidenced by the decline in the per capita murder rate, another source of the decline in the ratio between violent and non-violent crime arises from an increase in prosecutions for non-violent offences. The growth in this category of indictments is more rapid than even the most liberal estimates of population growth in London between 1760 and 1834, and cannot be attributed to this growth alone. In part, the increase could be due to a rise in actual crime, but is more likely to have resulted from the removal of capital punishment for most forms of theft between 1834 and 1842. This transition could potentially have encouraged victims to proceed with formal prosecutions, rather than relying on extra-judicial punishments and negotiations. An effect of this sort can be seen earlier in the data when, in 1808, Parliament removed the death penalty for pickpocketing, leading to a concomitant increase in prosecutions.[2] Meanwhile, the overall number of trials grew substantially, reflecting a criminal justice system responding to a rapidly growing population.

In 1834, the Old Bailey was given jurisdiction over the whole of England for serious crimes only; this leads to a systematic increase in the ratio of violent to non-violent crimes that can not be attributed to changes in behavior or legal culture. The fraction of violent crimes undergoes a sudden rise, driven in part by the increasing numbers of serious crimes committed outside of London that ended up in the Old Bailey, but more significantly by the decline in the number of prosecutions. The post-1840s decline in the number of non-violent crimes is cemented by the jurisdictional shift of the Criminal Justice Act in 1855 that gave greater scope to the Magistrates' courts. We examine the characteristics of this latter effect in the following section.

With these provisos in mind, the violent to non-violent trial ratio can be seen to divide in to three distinct phases: (1) slightly declining, from 1760 to 1834; (2) a transition in 1834 to about 1860 driven by jurisdictional change; (3) a second reasonably stable phase. The main underlying social processes that drove these dynamics include, for example, not just an overall decline in the murder rate, but also increasingly limited use of capital punishment; both of which are classic components of the civilizing process. Meanwhile, our statistical tests for a three-epoch model, described in the main text, allow us to reject a simple model where the JSD is driven by systematic effects associated with the indictment ratio (Fig. S6), or by the 1840s transition in the overall number of trials for violent indictments (Fig. S5).

S4: Indictments: Internal Structure

This section describes the changing composition of the Old Bailey indictments in greater detail, including relevant legal-historical details and cultural processes.

Our data contains a total of sixty distinct indictment types; the prevalence of which varies widely, both within a particular year, and over the course of the period reflected in the data. Because of the changing volume of trials over the course of the data, it is more revealing to consider the average yearly indictment fractions, rather than the overall fractions of indictments. Fig. S8 shows this distribution on a log-linear scale to give a sense of how much an "average" year at the Old Bailey concentrates on different types of crime.

A small number of indictments account for the majority of cases; the top eleven indictments account for 90% of all non-violent offences (the top five, in the case of violent indictments). Concern for property dominates both the violent and non-violent indictments; trials for violent theft ("highway robbery" and "robbery") average 54% of a year's docket for violent offences, while trials for theft (ranging from larceny to pocketpicking, poaching, and extortion) make up 84% of the non-violent docket. Other forms of violence—wounding, murder, manslaughter, infanticide and rape are the next five most common—are subdominant concerns; meanwhile, forms of deception not involving direct theft (counterfeiting, forgery, fraud, bigamy and perjury) are the next most common indictment types for non-violent offences, with other non-violent crimes of concern to the state (including non-violent sexual offences, return from transportation, libel, sedition, and illegal abortion).

The diversity of indictments—here quantified as the entropy of the distribution over indictments—is shown in Fig. S9. In contrast to the class of violent indictments, which maintains a reasonably constant diversity, the non-violent indictments are increasingly diverse over time. The transition

---

[2] 48 Geo. III, c. 129, s. 2 (1808); see Ref. [7].

is well-marked by the two major jurisdictional shifts; it begins around 1834, a period of definitional change, and finishes around 1855. Again, our statistical tests for a three-epoch model, allow us to reject a simple model where the JSD is driven by this two-phase diversity structure.

Fig. S10 shows the changing make-up of the list of non-violent indictments. The most visually striking shift, in 1827, is purely procedural in origin. Indictments for grand and petty larceny were combined by act of Parliament into "simple larceny".[3] One can also see the effect of the 1808 pickpocketing statute mentioned above. The upper red band corresponding to the pickpocketing indictments undergoes rapid expansion correlating with the implementation of new statute law. A similar pattern can be seen in the indictment class including "stealing from master", following an act in 1823 (upper blue band).

Fig. S11 shows the changing make-up of the list of violent indictments. The two most striking shifts are (1) the near-complete disappearance of the "highway robbery" violent theft category after 1829, and (2) the increasing representation of "wounding" as an indictment.

In a similar fashion to the consolidation of larceny indictments into a broader class, phenomenon (1) appears to stem from a purely formal change in legal categorization; this happened in two stages, with the first shift happening in 1827 (under the same acts consolidating larceny offences) and then in 1828;[4] these two acts together eliminated a range of previous indictment subcategories and established new definitions—a set of practices that took a year or so to propagate to the judges in the courtroom itself and leading to the recategorization of highway robbery under other forms of violent theft (here collected into the Old Bailey's violent theft/robbery category).

Phenomenon (2) appears as a much longer-term shift towards prosecutions for non-fatal assault. This appears to be a signature, at the crime-categorization level, of the same civilizing process that we detect at the fine-grained level of recorded speech. Meanwhile, despite the fact that the violent crime docket begins to include crimes of lesser severity (wounding as opposed to killing or violent theft), the overall distinguishability of violent and non-violent indictments continues to increase.

S5: The Civilizing Process by Indictment

Based on findings in the previous two sections, this section examines the civilizing process at the more fine-grained level of indictment classes. We find strong evidence for the same civilizing process found in the system as a whole, further confirming and elaborating the results of the main text.

In particular, we study the increasing trial distinctiveness (Jensen-Shannon Distance) between groups of indictments within each of the violent and non-violent classes. Because some indictments are very rare (see Fig. S8), we collect indictments into six major subgroups; three subgroups for violent indictments, and three for the non-violent indictments.

The non-violent subgroups are **Theft**, **Deception**, and **Other**.

In the OBO categorization system, **Theft** contains grandLarceny, pettyLarceny, simpleLarceny, theftFromPlace, burglary, pocketpicking, stealingFromMaster, receiving, animalTheft, housebreaking, shoplifting, embezzlement, mail, extortion, gameLawOffence, breakingIntoPlace and the "other" theft category; **Deception** contains fraud, forgery, perjury, bankruptcy, the "other" deception category and bigamy and coiningOffences; the **Other** category collects a number of offences that do not clearly fall under either property crime or simple deception, and includes arson, libel, returnFromTransportation, pervertingJustice, conspiracy, taxOffences, sodomy, keepingABrothel, concealingABirth, seditiousLibel, treason, seducingAllegiance, seditiousWords, religiousOffences, illegalAbortion, piracy, habitualCriminal, vagabond and "miscellaneous" other.

The violent subgroups are **Violent Theft**, (non-fatal) **Assault**, and **Killing**.

In the OBO categorization system, **Violent Theft** contains robbery, highwayRobbery, and the "other" violentTheft category. **Assault** contains wounding, assault, threateningBehaviour, rape, kidnapping, riot, indecentAssault and assaultWithSodomiticalIntent. **Killing** contains murder, manslaughter, infanticide, killing/pettyTreason, and the "other" killing category.

We can then ask (for example) whether **Violent Theft** and **Theft** show the same increasing distinctiveness found for the violent and non-violent categories as a whole. Table S1 shows the nine possible pairwise comparisons for both the coarse-grained and fine-grained semantic levels, and we show graphical results in Figs. S12 through S14.

Both **Violent Theft** and **Assault** show significant evidence ($> 5\sigma$) for a linear increase in trial distinguishability against all three non-violent indictment categories from 1778 to 1913. The most significant signal is the increasing distinctiveness of violent and non-violent theft, and the largest rate of increase is found for the distinction between assault and deception and violent theft and deception. Among all the pairs, violent theft and theft are the hardest to distinguish in Jensen-Shannon distance—as is to be expected, given that the entire distinction hinges on the involvement of violence—and this distinction undergoes significant growth over the course of the data.

Category (V)	Category (NV)	Slope (bits/decade)	std dev
Violent Theft	**Theft**	0.0011	+20.1
	Deception	0.0022	+14.6
	Other	0.0012	+8.5
Assault	**Theft**	6×10^{-4}	+6.2
	Deception	0.0026	+14.2
	Other	6×10^{-4}	+3.6
Killing	Theft	-3×10^{-4}	−3.1
	Deception	0.0013	+6.8
	Other	-1×10^{-4}	−0.8

Fig. S1. Linear model parameters and significance (standard deviations away from zero) for indictment class pairs at the "Category Level" coarse-graining (equivalent to top of Fig. 1 of the main text). Indictment group pairs with significant evidence of increasing distinctiveness ($> 5\sigma$) are marked in **bold**.

Category (NV)	Category (NV)	Slope (bits/decade)	std dev
Theft	**Deception**	-8×10^{-4}	−8.1
Theft	**Other**	-6×10^{-4}	−5.8
Deception	Other	$+6 \times 10^{-4}$	3.8

Fig. S2. Linear model parameters and significance (standard deviations away from zero) for indictment class pairs, at the "Category Level" coarse-graining, within the non-violent indictment class.

---

[3] 7 & 8 Geo. IV, c. 29: An Act for Consolidating and Amending the Laws of England Relative to Larceny and other Offences Connected Therewith

[4] 9 Geo. IV, c. 31: An Act for consolidating and Amending the Laws of England Relative to Offences Against the Person

Killing, by contrast, shows evidence for this process only in the case of deception. Comparison with Figs. S12 through S14 suggests that part of the reason is that strong distinctions *already* exist in 1778—it is only in the 1900s, for example, that trials for violent theft and trials for non-violent theft become as distinct as trials for theft and trials for killing. This suggests that what we are seeing in the years available to us here is the further "perfection" of a longer-running process.

This pattern is in contrast to the development of distinctions within the non-violent indictment class; as shown in Table S2 and Fig. S15, non-violent indictment classes themselves become less distinguishable over time.

S6: The Trial of Mary Hall

The Trial of Mary Hall for Grand Larceny, 28 October 1801.

MARY HALL was indicted for feloniously stealing, on the 19th of October, five guineas, the property of John Martin.

JOHN MARTIN sworn.—I am a labourer: On the 19th of October, I was going up Back-lane, Rosemary-lane, about two o'clock in the day; I met with a friend, and got a little too much; I was returning home, and met with the prisoner in Wellclose-square; I went home with her; she asked me what I would give her for a bed that night.

Q. What time was it?—A. I cannot exactly tell; I agreed to give her two shillings; we slept together till between five and six in the morning; I had my breeches under my head, and in them a purse, containing five guineas; when I waked in the morning, between five and six, I searched, and found my money safe; I got up to evacuate, and while I was doing that, she got the money from me; I saw her take the money out of my breeches; I then laid hold of her, and had a struggle with her, to get the money back again, and she cried out, and a man came into the room with a knife in his hand, and swore, if I did not stand back, he would run the knife into my heart; I still kept in contention with her, and then a woman, that he called his wife, came up with a pair of tongs, and hit me with the tongs on my back; I took the tongs from her, and then the man said, if I did not lay down the tongs, he would surely run the knife through my heart; I was afraid of being murdered, and I put on my breeches, my waistcoat, and my jacket, and there came a black man behind me, he struck me upon my back, and I made resistance till I got out into the street; then the people gathered about, and they threw my shoes and stockings out at window, and I put them on; I then went to Lambeth-street, and got an officer.

RICHARD OSMAN sworn.—I am an officer of Lambeth-street, Whitechapel: On Tuesday, the 20th of October, the prosecutor came to me, and said, he was robbed of five guineas; that he had been used very ill, and asked me to go with him to apprehend the man that had robbed him; he said, he saw the girl with the money in her hand; I went, and apprehended her; I searched her, and found upon her one shilling, and some halfpence; she said, she knew nothing about it, she had no other money; he told me, there was an old man drew a knife upon him.

Q. Who is that man?—A. It turned out to be a man of the name of Stinson; he saw him in the House of Correction; he told me exactly the same story before I apprehended the girl, that he has now told your Lordship.

Q. I suppose he was quite sober at that time?—A. Yes. Q. How came Stinson not to be committed?—A. Before the girl had her last examination, he was taken in execution, and sent to the House of Correction.

Prisoner's defence. The man came into the room to my assistance, with his breakfast in his hand.

GUILTY, aged 29.

Second Middlesex Jury, before Mr. Common Serjeant.

Total length: 563 words. Total spoken: 524 words. Total lexical semantics: 147 words. Old Bailey Online Trial ID t18011028-44; last accessed 23 November 2013.

Our classification of violent indictments, following OBO definitions: offenceCategory={kill or violentTheft} or offenceSubcategory={assault [assault, assaultWithIntent, assaultWithSodomiticalIntent and indecentAssault], riot, threateningBehaviour, rape, kidnapping, or wounding}; all other offence categories (including, for example, the grand larceny indictment above) are classed as "non-violent" indictments. A very small fraction of trials (181 out of 112,485) have multiple indictments for both violent and non-violent offenses; we classify these trials as "violent."

S7: Old Bailey Corpus vs. Google N-Grams

The Old Bailey Corpus is a particularly long-baseline sample of spoken English, produced in the particular context of the criminal justice bureaucracy. It is worth comparing the synonym set categorization of this data against a larger corpus, both to see large-scale trends in the English language as well as to check for the possibility of biases that might exist due to the changing relevance of Roget's categorization. The top twenty-four most relevant synonym sets for our analysis are shown at the end of this section. There we compare the synonym-set frequency in the OBC to that found for the top ten words of those sets in the Google N-Grams corpus (British English, Ref. [8]). Since the Google N-Grams corpus normalizes by all words (as opposed to by semantically significant words), frequencies are roughly a factor of five lower; we scale the Google corpus by this factor to make visual comparison easier.

As can be read from those figures, most of the variation in the Old Bailey Corpus is only weakly correlated to variation in the use of British English in the Google corpus as a whole (Pearson cross-correlation, 0.10 ± 0.02). Further, in most cases, the synonym set density in the Google corpus shows significantly less relative variance than in the OBC (11% *vs.* 29%, $\pm 1\%$), suggesting that the bureaucratic system itself is undergoing greater change than the culture as a whole.
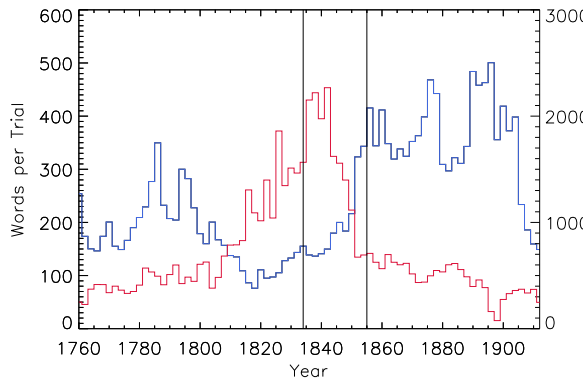
Fig. S3. Semantically significant words per trial (blue line) and total number of trials per year (red line) in the OBC.
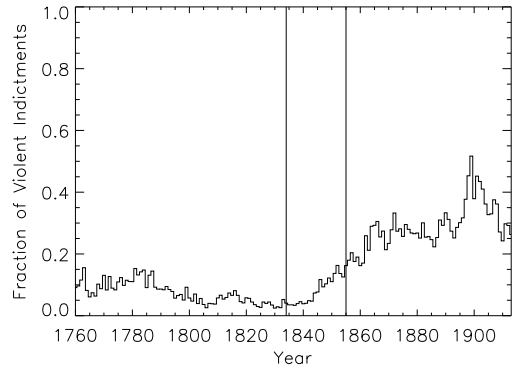

Fig. S6. Overall ratio of trial indictments, violent to non-violent, in the data.
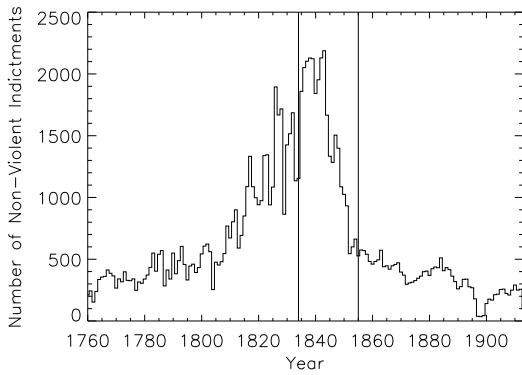

Fig. S4. Total number of indictments for non-violent crimes in the corpus.
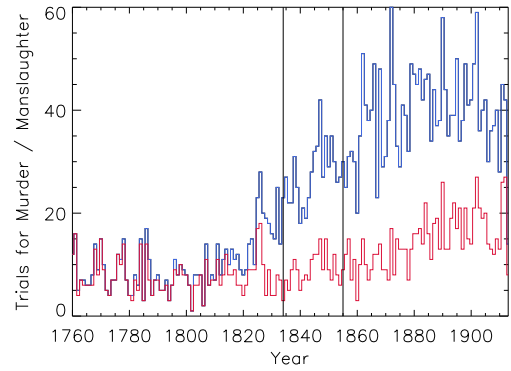

Fig. S7. Total number of trials in the OBO (not OBC) for murder and manslaughter (blue line) and murder alone (red alone).
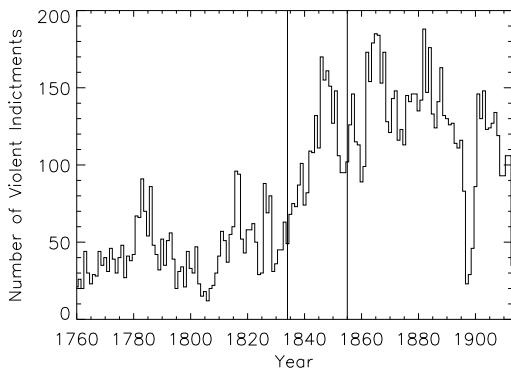

Fig. S5. Total number of indictments for violent crimes in the corpus.
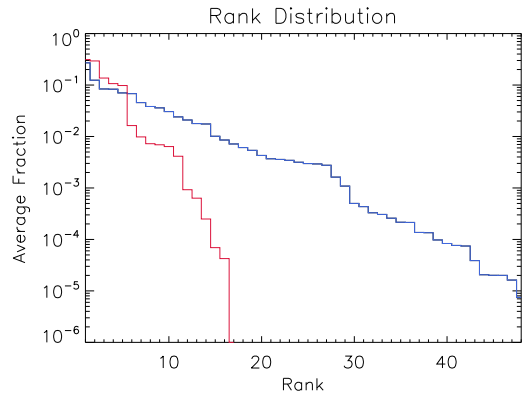
Fig. S8. Distribution of offence category-subcategory pairs within the OBC; broken out by indictments for non-violent offences (48 indictments total; blue line) and violent offences (16 total; red line).
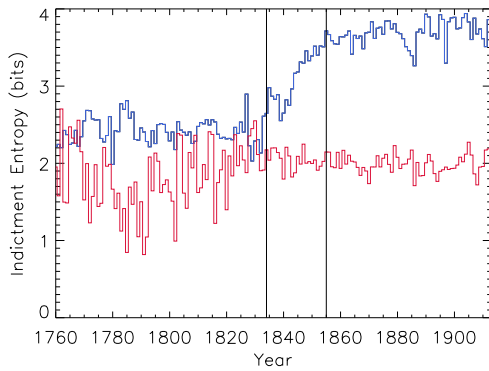
Fig.   S9. Diversity of indictment types, broken out into violent (red line) and non-violent (blue line) categories. While the entropy for the violent indictment class remains roughly constant over 150 years, the non-violent indictments become increasingly diverse over time, undergoing a distinct phase shift between 1830 and 1850.
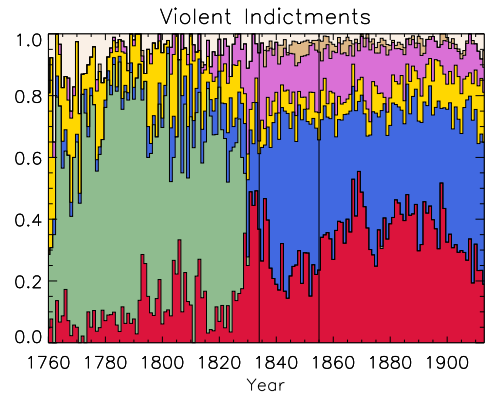
Fig.   S11. Composition of the violent indictment class. From bottom to top, violent theft/robbery (red), violent theft/highway robbery (green), wounding (blue), murder (yellow), manslaughter (purple), simple assault (brown); the remainder in white.
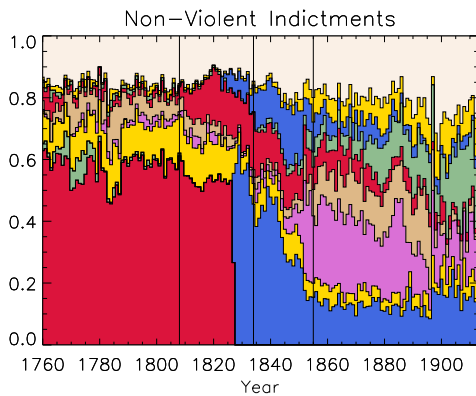
Fig.   S10. Composition of the non-violent indictment class. From bottom to top, grand larceny (red), petty larceny (green), simple larceny (blue), theft from place (yellow), coining offences (purple), burglary (brown), pocketpicking (red), fraud (green), "stealing from master" (blue), forgery (yellow), receiving stolen goods (purple); remainder in white. Simple larceny combined the grand and petty larceny cases by statue in 1827. In addition to the 1834 and 1855 points, we overlay the 1808 date, when pocketpicking was removed from the list of capital offences.
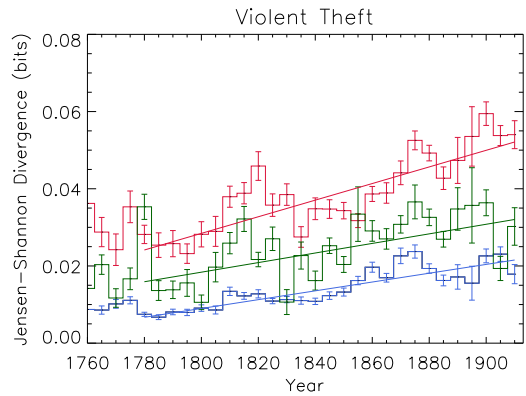
Fig.   S12. The Jensen-Shannon Divergence for violent theft and (1) theft (blue line), (2) deception (red line), and (3) all other non-violent offences (green line). In all three cases, strong evidence exists for the same civilizing process observed in the overall evolution of the Old Bailey. As expected, the hardest to distinguish pair is violent theft and theft (the JSD is lower than the other two pairs), but still shows significant ($p \ll 10^{-5}$) evidence for increasing divergence over time.
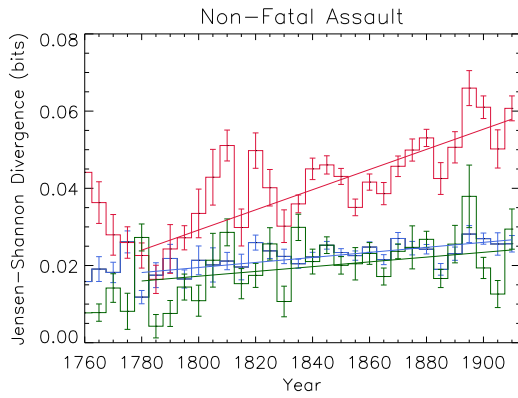
Fig. S13. The Jensen-Shannon Divergence for non-fatal assault and (1) theft (blue line), (2) deception (red line), and (3) all other non-violent offences (green line). Strong evidence exists for an increasing divergence over time between both assault and theft, and assault and deception.
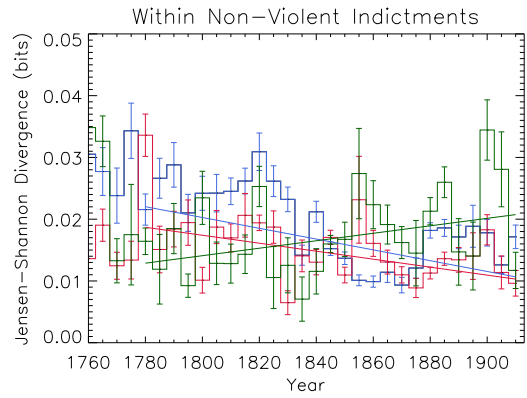
Fig. S15. The Jensen-Shannon Divergence for indictment types within the non-violent class: (1) theft–deception (blue line), (2) theft–other (red line), and (3) deception–other (green line). In contrast to the violent–non-violent genre emergence, deception and "other" become increasingly *less* distinct from the larger theft class over time, and only marginal ($< 4\sigma$) evidence exists for an increasing distinction between deception and other.
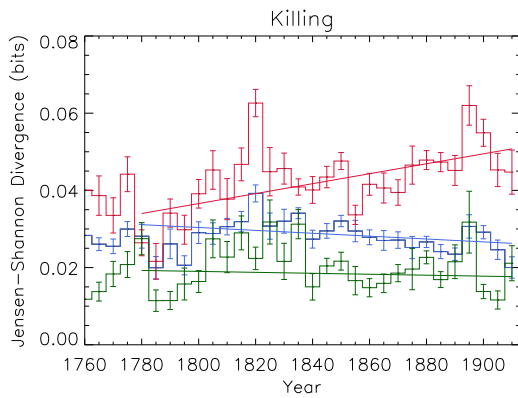
Fig. S14. The Jensen-Shannon Divergence for offences involving killing (murder, manslaughter, infanticide, treasonous killing) and (1) theft (blue line), (2) deception (red line), and (3) all other non-violent offences (green line). Here, only deception and killing show strong evidence ($> 5\sigma$) for increasing divergence over time.
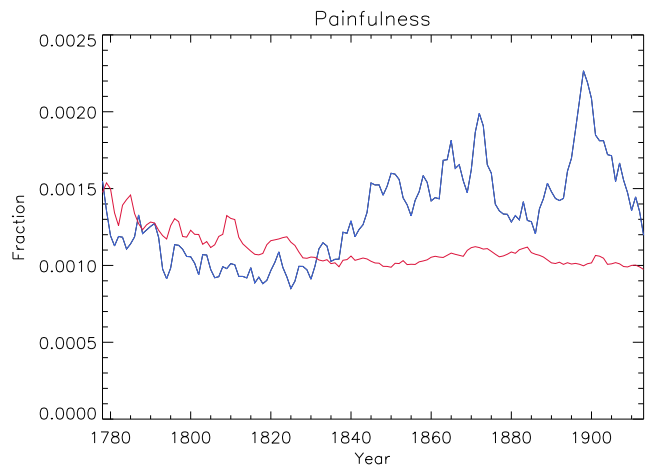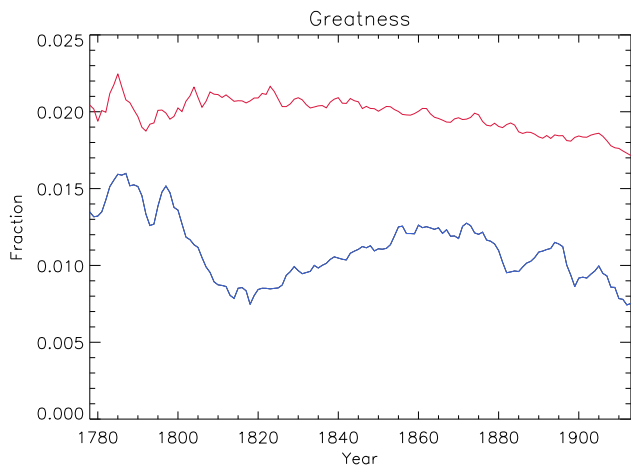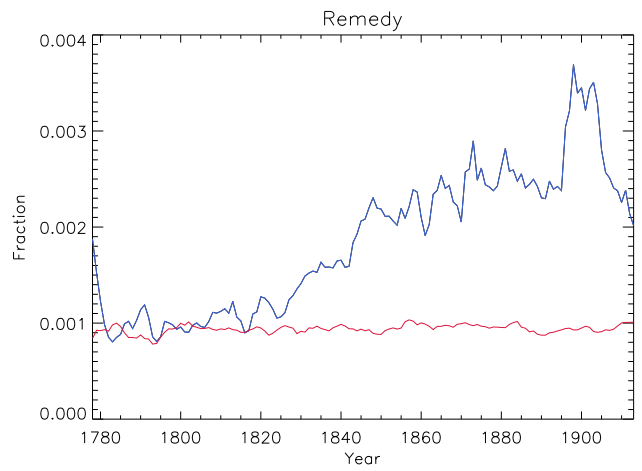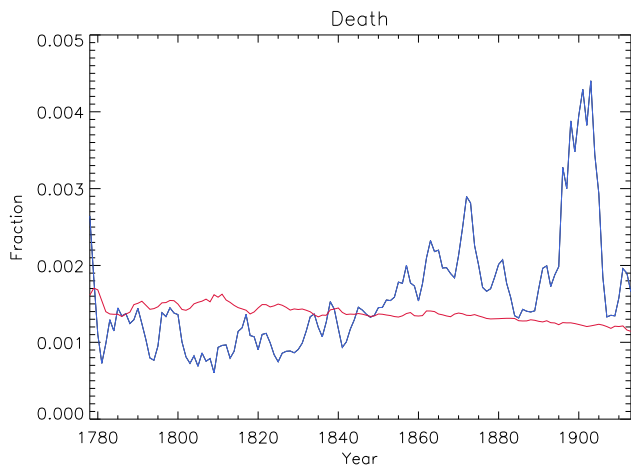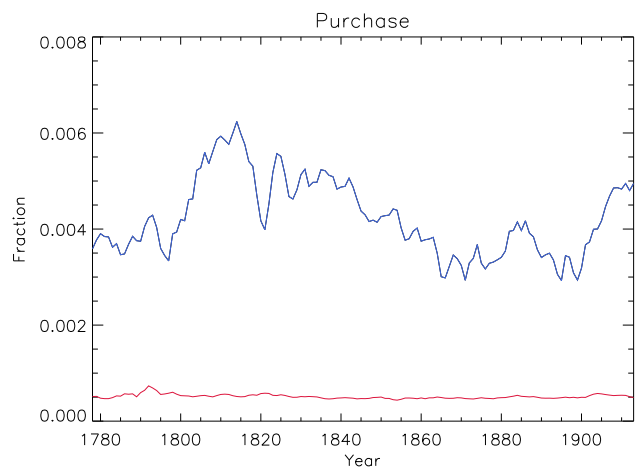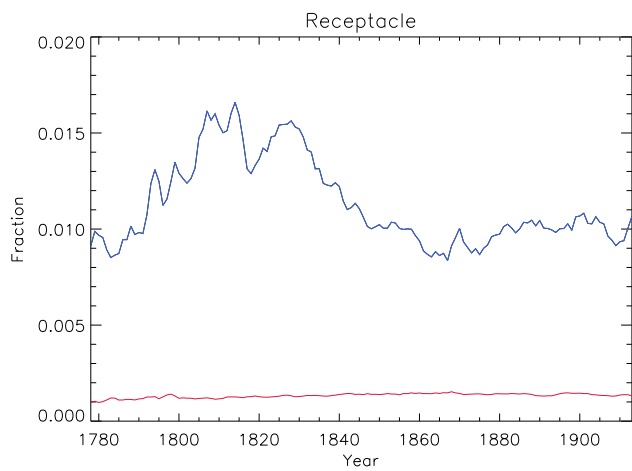
Fig. S16. Overall abundance of words. Old Bailey Corpus (blue line) *vs.* Google N-Grams British English database (red line, $\times 5$) for the top synonym sets associated with violent offences, Fig. 3(a) of the main text.
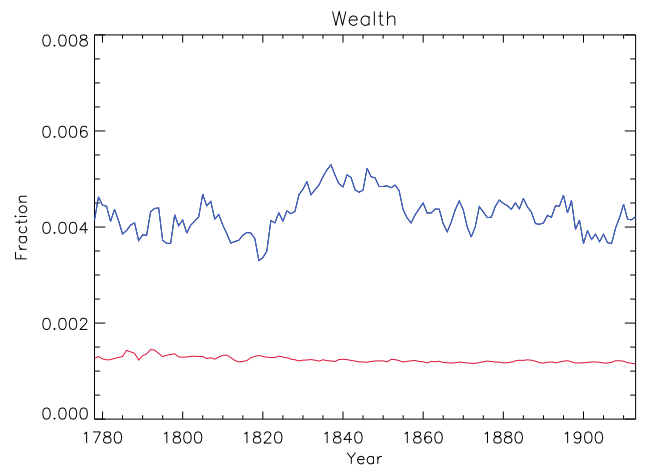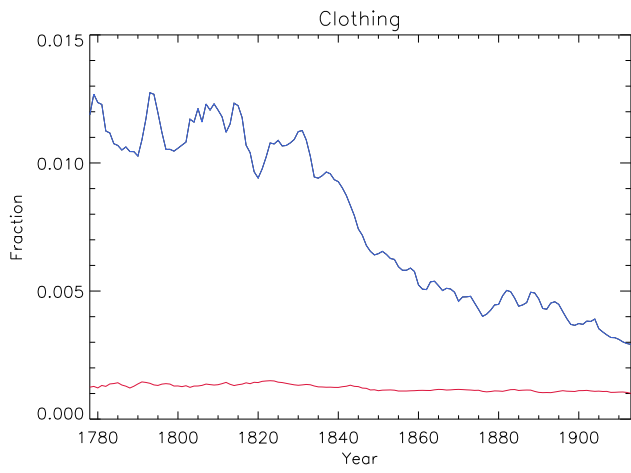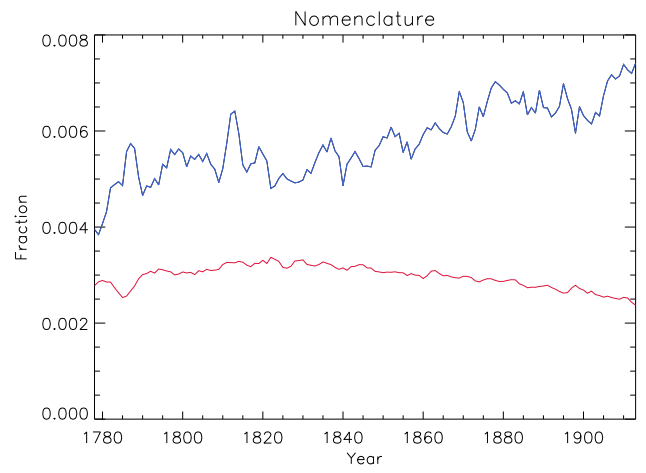
Fig. S17. (continued).
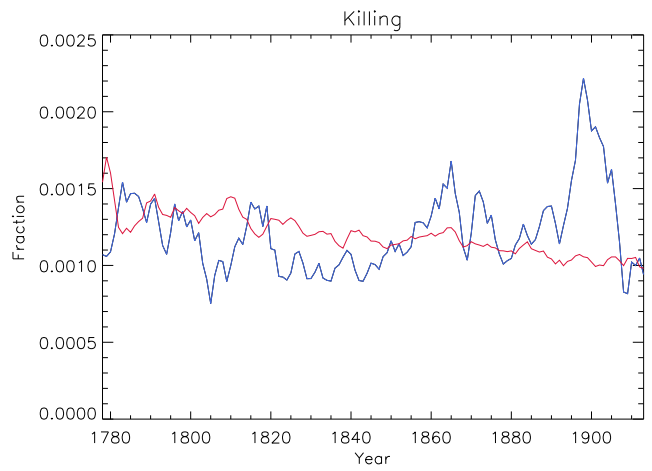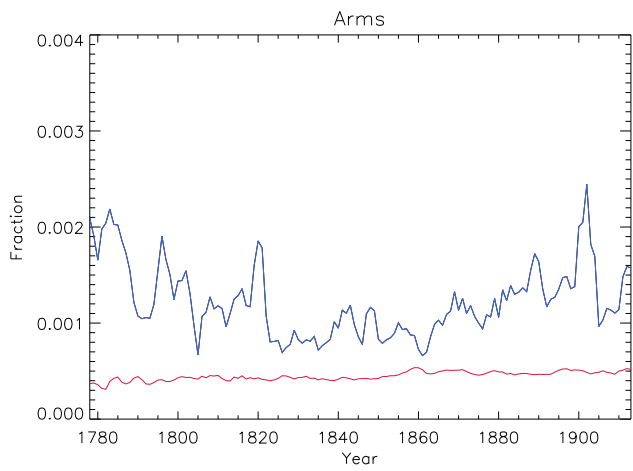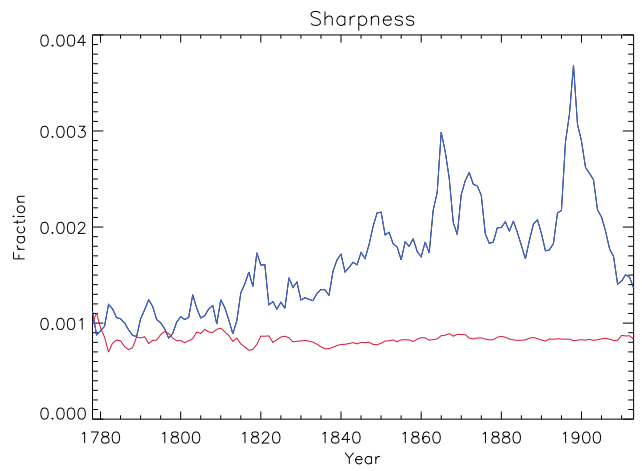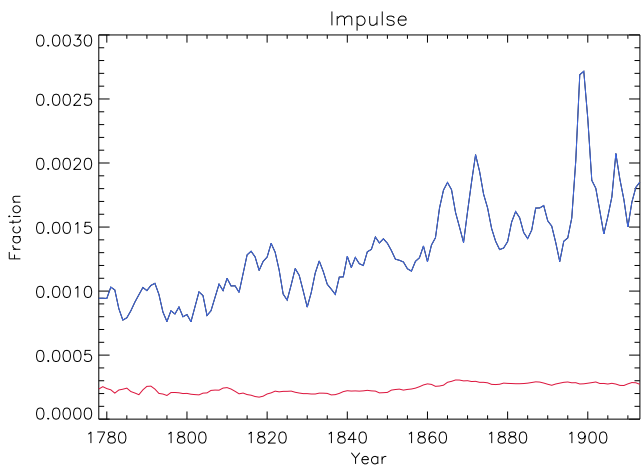
Fig. S18. (continued).

Fig. S19. (continued).
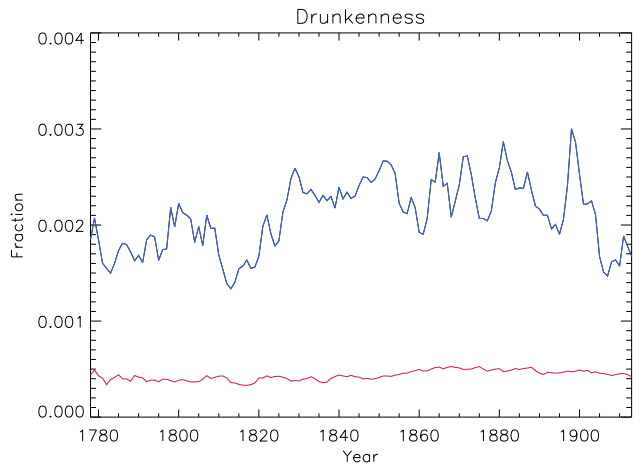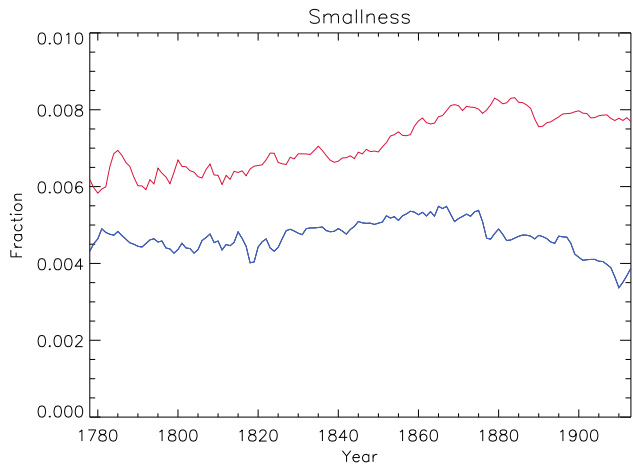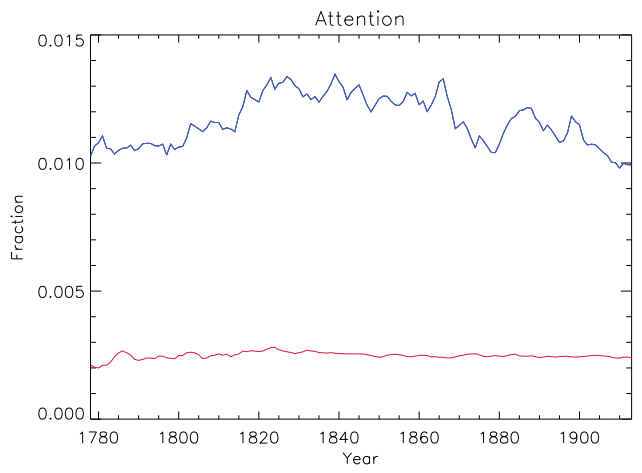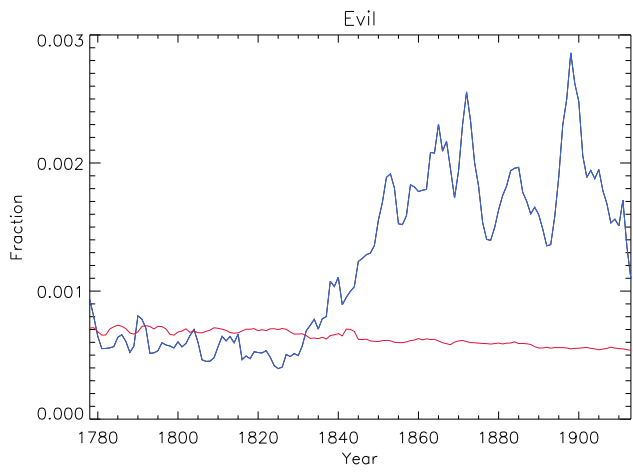
Fig. S20. (continued).

Fig. S21. (continued).
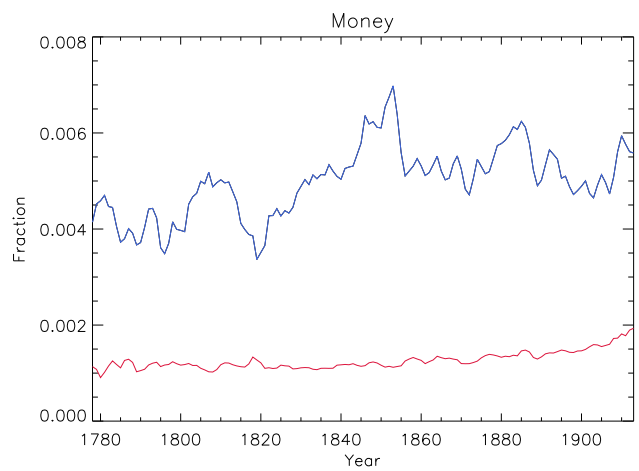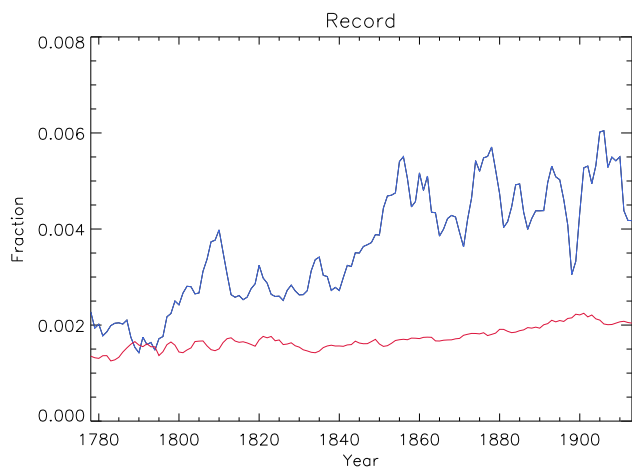
Fig. S22. Overall abundance of words. Old Bailey Corpus (blue line) *vs.* Google N-Grams British English database (red line, ×5) for the top six synonym sets associated with non-violent offences, Fig. 3(b) of the main text.
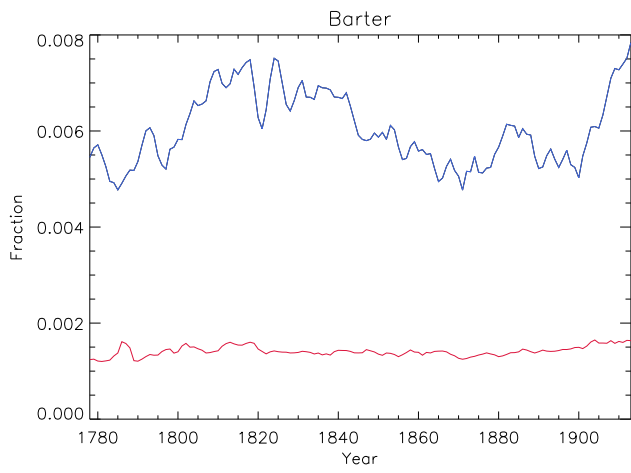
Fig. S23. (continued).

Barter

Nomenclature

Clothing

Wealth

Fig. S24. (continued).
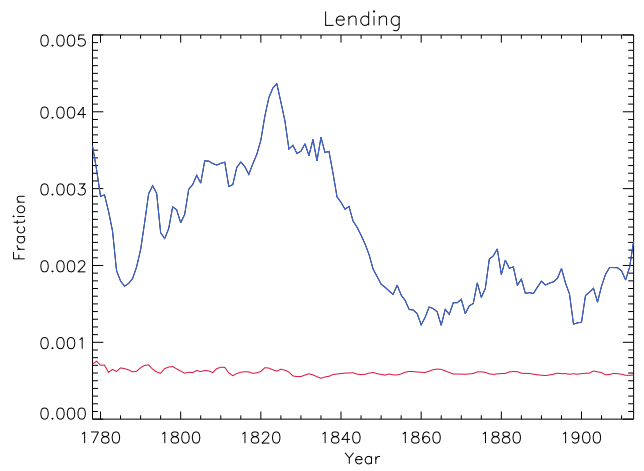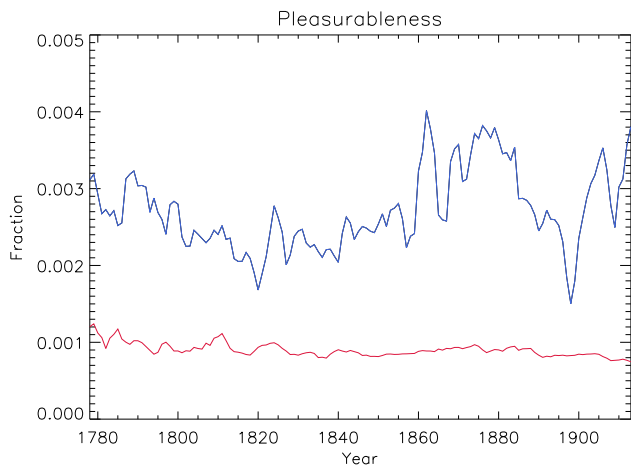
Fig. S25. (continued).
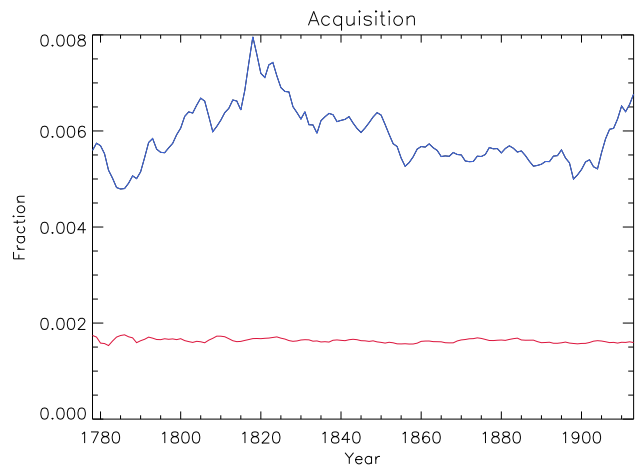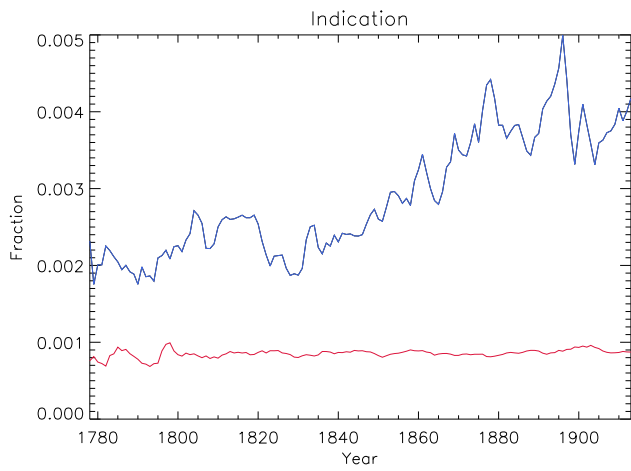
Fig. S26. (continued).

Fig. S27. (continued).

1. Emsley C (2010) Crime and society in England, 1750–1900 (Longman, London) Fourth edition, Ch. 8.

2. Williams CA (2000) Counting crimes or counting people: some implications of mid-nineteenth century british police returns. Crime, Histoire & Sociétés 4:77–93.

3. DeDeo S, Hawkins RX, Klingenstein S, Hitchcock T (2013) Bootstrap methods for the empirical study of decision-making and information flows in social systems. Entropy 15:2246–2276.

4. Eisner M (2003) Long-term historical trends in violent crime. Crime and Justice 30:83–142.

5. Cockburn JS (1991) Patterns of Violence in English Society: Homicide in Kent 1560-1985. Past & Present 130:70–106.

6. Gurr TR (1981) Historical Trends in Violent Crime: A Critical Review of the Evidence. Crime and Justice 3:295–353.

7. Beattie J (2012) The First English Detectives: The Bow Street Runners and the Policing of London, 1750-1840 (OUP Oxford) Pg. 211.

8. Michel JB, et al. (2011) Quantitative Analysis of Culture using Millions of Digitized Books. Science 331:176–182 British English Corpus; actual search at `http://bit.ly/1e5AvqY`; last accessed 11 December 2013.