Crime and Prejudice: The Use of Character Evidence in Criminal Trials

Joel Schrag
Emory University

Suzanne Scotchmer
University of California, Berkeley

The task of juries is to dispense ex post justice. While justice requires convicting the guilty and acquitting the innocent, the evidence usually cannot distinguish with certainty. We argue that the jury will be more lenient in acquittals than is optimal for deterring crime whenever its subjective cost of wrongful convictions is at least as high as its subjective cost of wrongful acquittals. However if the jury is prejudiced against habitual criminals, its subjective cost of wrongful convictions will be relatively low, and then the jury may impede deterrence by its punitiveness rather than by its lenience. We investigate whether restricting character evidence can solve this problem.

1. Introduction and Summary
Criminal justice is meted out in a hierarchy that includes police, prosecutors, judges, and juries. These actors are motivated and constrained in different ways. To a first approximation, police want to solve crimes, prosecutors want to secure convictions, and the jury wants to convict the guilty. While there is a certain coherence to this decentralized system, no single actor coordinates the system to achieve broad social goals such as deterrence. In this article we expose a potential conflict between the social goal of ex ante deterrence and the judicial goal of ex post justice, and ask whether that conflict can be mediated by rules of evidence—in particular the rule that excludes character evidence from criminal trials.

For maximal ex ante deterrence the standard of evidence should maximize the difference in expected costs suffered by the citizen when guilty and when innocent. There are two stages at which the judicial hierarchy treats guilty citizens different from innocent ones. First, the probability of being charged with a crime presumably depends on whether or not the citizen is guilty. Second, the probability of being convicted depends on the evidence the citizen

---

We thank Eddie Dekel, Gillian Hadfield, Matthew Rabin, Kate Rockett, Susan Rose-Ackerman, and Dan Rubinfeld for useful discussions. Alan Schwartz and two anonymous referees provided additional useful comments. This research was supported by the National Science Foundation, Grant SES 90 119 10, the John M. Olin Foundation through grants to Boalt School of Law and Yale Law School, and the Alfred P. Sloan Foundation.

© 1994 by Oxford University Press. All rights reserved. 8756–6222/94/$5.00
generates, which is related to whether or not the citizen is guilty. The first stage is the responsibility of police and prosecutors, while the second belongs to the judge and the jury.

An example that starkly exposes the potential conflict between ex ante deterrence and ex post justice is as follows. Suppose there is only one citizen who could possibly commit a particular crime; that is, the crime opportunity is “exclusive.” If the crime occurs and the potential criminal is charged, then that citizen should be convicted; this is obvious. What is perhaps less obvious is that for maximal deterrence, any defendant should be convicted, and the argument does not require that the police or jury know which citizens had which crime opportunities. They need to know only that the crime opportunity was exclusive. If the police arrest the right person, then justice is done by convicting, and the prospect of a conviction might deter the potential criminal. But the police might instead arrest an innocent person. Even if innocent persons anticipated that they might be convicted of such a crime, this would not affect their behavior because that particular crime was not in their opportunity set and they did not make a decision about it. From the innocent person’s point of view, being convicted of the crime is random bad luck, but there is nothing the innocent could have done to prevent it. Of course, in the interest of justice we would not want to convict every defendant who comes to trial. For justice the jury will use the evidence at trial to distinguish guilt from innocence, even though for deterrence the evidence should be ignored.

The foregoing example suggests that juries will be too lenient relative to maximal deterrence, since they want to protect innocent defendants in a way that has no bearing on deterrence. Another source of leniency is the judge’s instructions to the jury to behave as if the indictment contains no information about guilt or innocence. If the jurors followed these instructions they would “believe” prior to seeing any other evidence that the accused is no more likely to be guilty than a person randomly pulled off the street—that is, if the population were size $N$, then the prior probability of guilt would be $1/N$. With such a prior probability, the jury would convict only on very strong evidence, which may be lenient relative to the standard of evidence that would maximize deterrence. It might seem that this leniency could be overcome if the jury incorporated its knowledge of the indictment process into its beliefs. Our first result below, however, shows that if the jury’s perceived cost of wrongful convictions is at least as great as its perceived cost of wrongful acquittals, then the jury will always be too lenient relative to maximal deterrence, even if the jury rationally updates its beliefs to reflect the information contained in the indictment.

When there is a conflict between ex ante deterrence and ex post justice, it is hard to know how to balance one against the other. There is no social overseer to resolve the conflict. To some extent judges try to modify judicial outcomes by telling the jury what kind of evidence is sufficient for conviction, but their power in this domain is limited. Judges may also try to modify judicial outcomes by telling jurors to ignore certain kinds of information, such as the
information contained in the fact that the defendant was arrested and indicted. We assume that such information is impossible to ignore, even if that would be desirable. Our formal assumption is that juries are rational in the sense of updating their information in a Bayesian way, using all their knowledge of the judicial process and evidence. A third type of intervention, the one on which we focus here, is made through judicial rules of evidence, which restrict what the jury can know about the crime and about the defendant.

The Federal Rules of Evidence (FRE) direct the judge to balance the probative value of evidence—that is, how much it affects the probability of guilt versus innocence—against its potential to prejudice the jury against the defendant. Evidence “may be excluded if its probative value is substantially outweighed by the danger of unfair prejudice. . . .”¹ In particular, Rule 404 prohibits the use of character evidence, including conviction records, to demonstrate that a defendant is predisposed to commit illegal acts and is therefore likely to have committed the crime for which he or she is charged.² Rule 404 addresses the fear that a jury may believe that it is “just” to convict a defendant on the basis of “bad character” even if the rest of the evidence is weak. Dean Wigmore (1935) points out that such a view is not unnatural and even appears in literature. In punishing Tom Sawyer for a transgression she could not prove, his aunt says, “Well, you didn’t get a licking amiss, I reckon. You’ve been into some other audacious mischief when I wasn’t around, like enough.” (Twain, 1875: 40).

The Rules of Evidence frame the debate about character evidence as a question of whether it is probative or prejudicial. Evidence has probative value if it has a “tendency to make the existence of any fact . . . more probable or less probable than it would be without the evidence.”³ In statisticians’ terms, evidence is probative if it changes the jurors’ beliefs about the likelihood of the defendant’s guilt. Evidence is prejudicial if it makes the fact finder “view the defendant as deserving of punishment . . . and . . . [be] willing to convict even if there were a reasonable doubt about the defendant’s guilt” (Kuhns, 1981: 778) or, alternatively, if it evokes “punitive instincts . . . [,] a dark psychological force that impedes the effective operation of probative light” (Weissenberger, 1985: 585).

There is an obvious rationale behind the probative/prejudicial distinction: If the jury pursued the “right” social objective, then society’s outcome could only be improved by improving the jury’s information, that is, by giving it probative information. And we would not want to distort that outcome by giving the jury information that makes it irrationally prejudiced.

---

2. FRE 404 (a)-(b) states: “Evidence of a person’s character or a trait of character is not admissible for the purpose of proving action in conformity therewith on a particular occasion . . . Evidence of other crimes, wrongs, or acts is not admissible to prove the character of a person in order to show action in conformity therewith.” (Mueller and Kirkpatrick, 1990: 39).
However, our point in this article is that the jury does not represent all of society's objectives; it undervalues deterrence. Even if evidence is probative, it might cause an already too lenient jury to be more lenient, and even if the evidence is prejudicial, permitting such evidence might serve the public interest by leading to convictions that would otherwise not be made. We therefore put aside the probative/prejudicial test for admissibility, and go directly to the task of trying to assess how admitting or excluding character evidence affects deterrence when the jury is concerned only with ex post justice.

We take character evidence to be probative if it changes the jury's posterior probability of guilt in a way that rationally reflects the judicial process and equilibrium crime rates. We take it to be prejudicial if it changes the jury's inclination to convict without changing the probabilities of guilt and innocence. Our model permits the police to be prejudiced as well as the jury. We say the police are prejudiced if, everything else being equal, they are more inclined to charge habitual criminals with crimes than noncriminals—that is, they "bring in the usual suspects."

In this environment, restricting character evidence affects judicial outcomes in at least two ways. First, it prevents the jury from updating its posterior probabilities of guilt versus innocence according to whether the police treat habitual criminals differently from noncriminals and according to the crime rate of the defendant. Second, it prevents the jury from obeying its own prejudicial whims, to the extent that those reflect the defendant's character.

As a benchmark, we first ask how restrictions on character evidence affect crime when neither the police nor jury are prejudiced. We suppose that there are two types of citizens: those with high valuations of crime and those with low valuations. In equilibrium, high-valuation citizens have higher crime rates than low-valuation citizens. Consequently, when the jury uses character evidence to conclude that the defendant has a high valuation of crime, it assesses a higher prior probability of guilt at trial and sets a lower standard of evidence. Now suppose we disallow character evidence. The jury then cannot choose different standards of evidence for high-valuation (high-crime) and low-valuation (low-crime) defendants. As a consequence, their crime rates will diverge even further, since the jury becomes more lenient toward habitual criminals and less lenient toward noncriminals. The effect on overall crime is ambiguous, but if crime opportunities are sufficiently overlapping, we show that admitting character evidence reduces crime.

Now suppose the police, but not the jury, are prejudiced. In particular, suppose the police are inclined to "bring in the usual suspects," namely, citizens with criminal records. Prejudice on the part of the police can lead to a self-sustaining equilibrium with two crime rates even if all citizens start out with the same valuation of crime. Citizens with criminal records are more likely than others to be indicted for crimes they didn't commit, and the jury realizes this. Consequently the jury will assign a lower probability of guilt, conditional on the evidence, to a defendant who is a habitual criminal and will be less willing to convict. The jury's leniency reinforces the habitual criminal in pursuing a life of crime. If we deny the jury evidence on character, the jury
can no longer "undo" the police prejudice, and will use the same standard of evidence for all defendants. In equilibrium all citizens with the same valuation of crime will end up with the same crime rate. Instead of putting a further wedge between the crime rates of habitual criminals and noncriminals, as above, it brings their crime rates closer together, also with an ambiguous effect on total crime, although in this case we show that if crime opportunities are sufficiently overlapping, admitting character evidence increases crime.

Finally we come to jury prejudice. By jury prejudice we mean that for a given posterior probability of guilt the jury is more willing to convict a habitual criminal than a noncriminal. Even if the jury thinks the defendant is probably not correctly matched to the crime at hand, it is willing to convict because many of the criminal's other transgressions probably will go unpunished. While such reasoning might promote justice, we argue that it hinders deterrence.

We show that allowing character evidence to be considered can make the jury too punitive toward habitual criminals, and that withholding character evidence can improve deterrence by making the jury more lenient. This should seem counterintuitive to the reader, since we have previously argued that juries are typically too lenient and that deterrence could in general be improved by making them more punitive. But the conclusion that juries are too lenient relies on an assumption that is overturned by the presence of prejudice. The assumption is that juries mainly seek to avoid wrongful convictions since the implicit cost they assign to wrongful convictions is at least as large as the implicit cost they assign to wrongful acquittals. In contrast, prejudice means the jury feels very little cost from wrongfully convicting a habitual criminal and considerable remorse from a wrongful acquittal. Hence the jury will convict on weak evidence.

In the extreme case, there may be a self-reinforcing equilibrium where juries are very willing to convict habitual criminals, whatever the evidence matching the criminal to the crime, and the police are very willing to arrest them. Since the criminal saves nothing by avoiding crime, there is no deterrence. In equilibrium the punitive police and jury reinforce criminal behavior, and criminal behavior reinforces their inclination to be punitive.

Such an equilibrium can be avoided by disallowing character evidence. If the jury does not know whether the defendant is a habitual criminal, its inclination to be punitive is reduced. The jury's common standard of evidence for all defendants will be higher than the one it would use for habitual criminals but lower than the one it would use for normally law-abiding citizens. Since the jury was previously too lenient with ordinary citizens (according to our previous argument), such a change can decrease the crime rate of both types.

Is this scenario likely? It depends not only on the jury's prejudice, but also on whether opportunities for crime are exclusive. As we argued above, when opportunities for crime are exclusive, deterrence is improved by a punitive jury, and that conclusion holds even if the jury's punitiveness is due to prejudice. It is only when several citizens might share the same crime opportunity
that a jury might be "too punitive." In that case, a citizen might be charged for a crime he or she could have done but did not do. If the police are imperfect at matching the perpetrator to the crime, and if the jury is willing to convict on weak evidence, then there is no incentive for the potential criminal to be innocent rather than guilty. If the citizen forgoes committing the crime, someone else might commit it instead, and the abstainer's probability of being charged and convicted is not significantly lower than if he or she had in fact committed the crime. A less punitive jury would require stronger evidence against the defendant to convict, and since stronger evidence is more likely if the defendant is guilty, this would improve deterrence.

Since some of these arguments turn on whether or not crime opportunities are exclusive, it is worth developing the concept of a crime opportunity. A crime opportunity available to an accountant, for example, is the embezzlement of company funds. Whether the opportunity is exclusive depends on whether other employees also have access to the funds. An opportunity for crime arises on the street when a potential mugger and potential mugging victim find themselves face to face. Whether the crime opportunity is exclusive depends on whether several potential muggers are present. In neighborhoods where it is hard to step onto the street without facing potential muggers, crime opportunities are nonexclusive. And indeed this is the context that springs to mind when one thinks of police "bringing in the usual suspects" and juries being aware of criminal histories. One can easily imagine (and perhaps observe) the pathological equilibrium where street criminals are reinforced in crime by indiscriminate police and juries, and police and juries are reinforced in their willingness to be punitive by the defendants' long histories of convictions.

For simplicity, we suppose that the crime opportunities available to potential criminals are exogenous. In reality, of course, the number of crime opportunities is affected by the behavior of potential victims, which in turn depends on criminals' behavior. Thus, the observed amount of crime is determined by an equilibrium between victims and criminals. For a discussion of this phenomenon, see Cook (1986).

In this article we maintain the central hypothesis of other economic literature on criminal deterrence, such as Becker (1968), that crime is the consequence of individuals' rational responses to the incentives that they face. For a review of this literature, see Cook (1980). We take this literature in a new direction by focusing on how the criminal justice system makes decisions and the consequent effect on potential criminals' incentives.

In Section 2 we develop the idea that juries will in general be too lenient relative to deterrence, and in Section 3 we develop our ideas about the consequences of restricting character evidence.

2. On the Leniency of Juries

We must model the behavior of both citizens and juries. We assume the citizen takes any crime opportunity whose value exceeds the expected cost. Should the authorities indict a citizen for a crime, the jury examines the evidence and
decides whether to convict or acquit. We assume that the jury is concerned
with ex post justice (and not explicitly with ex ante deterrence), and we
formalize this notion by assuming that the jury minimizes the expected costs
of Type I errors (convicting the innocent) and Type II errors (acquitting the
guilty).

The strength of the evidence presented at trial, \( e \), is a random variable
distributed on an interval \([0,\infty)\) according to probability density functions
\( f(e|i) \) and \( f(e|g) \), where \( i \) means the citizen was innocent of the crime for
which he or she was charged and \( g \) means guilty. A low value of \( e \) corresponds to
weak evidence matching the defendant to the crime. Thus we assume that an
individual may generate strong circumstantial evidence despite actual inno-
cence. As we discuss below, the jury chooses a standard of evidence, namely
the minimum evidence \( e \) sufficient for conviction.

We assume that the distributions of evidence satisfy the monotone likeli-
hood ratio property—that is, that the ratio of densities \( f(e|g)/f(e|i) \) increases
with \( e \). Stronger evidence \( e \) shifts the jury’s posterior probability away from
innocence toward guilt. This assumption implies that \( F(e|i) > F(e|g) \) for all \( e \),
where \( F(e|i) \) and \( F(e|g) \) are the cumulative distribution functions.

We will think of citizens as having “opportunities” for crime. Crime oppor-
tunities have benefits \( m \geq 0 \) distributed according to the density function \( c(m) \)
with support on an interval. The associated cumulative distribution function is
\( C(m) \). A crime may or may not be in a citizen’s “opportunity set,” but if it is,
the citizen must decide whether or not to take it. When deciding whether or
not to take a crime opportunity, a citizen compares the benefit from commit-
ting it with the expected increase in his or her punishment costs. The latter
depends on both the probability of the citizen being investigated, indicted,
and convicted if guilty and the probability of being investigated, indicted, and
convicted if innocent.

We assume that a citizen who commits a particular crime will be a suspect
(investigated by the police) with probability \( q \). If the citizen does not commit
a crime of value \( m \) in his or her opportunity set, then someone else commits it
with probability \( \mu(m) \), \( 1 \geq \mu(m) \geq 0 \). If someone else commits a crime in a
citizen’s opportunity set, the innocent person will still be a suspect with
probability \( p \). Thus, a citizen who forgoes a crime in his or her opportunity set
still becomes a suspect for that crime with probability \( \mu(m)p \). The probability
\( \mu(m)p \) depends on the nature of the crime, as well as on equilibrium behavior
of citizens. If the crime is the “exclusive” opportunity of a particular citizen,
then \( \mu(m) = 0 \), while \( \mu(m) \) can be positive if people share crime opportunities.

We assume that the police might erroneously arrest a citizen who did not
have opportunity to do the crime, and that the probability that a person will
become a suspect for a crime outside his or her opportunity set is \( p' \). We
assume that \( q > p, p' \).

After a citizen becomes a suspect, the prosecutor examines the evidence
and decides whether to press charges. If we assume that the prosecutor ob-
serves the evidence \( e \) and sends a case to trial if and only if the defendant
would be convicted, then our conclusions about what standard of evidence the
jury will choose and what crimes a citizen wants to commit are the same as if the prosecutor sends all cases to trial. For simpler notation we will therefore assume that all citizens investigated by the police are sent to trial. In reality, the prosecutor probably only observes a noisy signal of \( e \), but our conclusions are also robust to that modification.

The citizen has a value \( M > 0 \) of engaging in law-abiding activities, so \( M \) is the opportunity cost of crime. \( M \) is determined by the citizen's tastes and the value of the citizen's time spent in law-abiding activities. A citizen is willing to take a crime opportunity with value \( m \) if the benefit, \( m - M \), exceeds the expected increase in punishment costs, which we represent by the function 

\[
K(e, \mu, p, q) = \{q[1 - F(e|g)] - \mu p[1 - F(e|i)]\}S.
\]

\( S \) is the monetized value of the sentence that a citizen will receive if convicted, and \( [1 - F(e|x)] \) is the probability of being convicted when the jury's standard of evidence is \( e \) and the citizen takes action \( x \in \{i[\text{innocent}], g[\text{guilty}]\} \). Crimes that are not in a citizen's opportunity set do not enter that person's decision problem even though he or she might be indicted for such crimes. Consequently, the probability \( p' \) does not enter the citizen's problem. Thus the citizen will take all crime opportunities that satisfy the following condition:

\[
\pi(m, e, \mu) = m - M - K(e, \mu, p, q) \geq 0,
\]

where \( \pi \) represents the citizen's net payoff from taking a crime opportunity.

Our objective is to characterize the citizens' crime rate \( \sigma(e) \), and to that end we must characterize their best-reply "cutoff" \( m(e) \) such that citizens commit crimes \( m \geq m(e) \) but not crimes \( m < m(e) \). When crimes are "exclusive," \( \mu(m) = 0 \) for all \( m \), and \( m(e) \) solves \( \pi(m(e), e, 0) = 0 \). If citizens share crime opportunities, and if each chooses the same \( m(e) \), then \( \mu(m) = 1 \) for \( m \geq m(e) \) and \( \mu(m) = 0 \) for \( m < m(e) \). The citizens' best reply \( m(e) \) maximizes \( \int_{m(e)}^{\infty} \pi(m, e, 1) dC(m) \), and satisfies \( \pi(m(e), e, 1) = 0 \) or \( m(e) = M + K(e, p, q) \). We assume that crime opportunities are distributed randomly among the \( N \) citizens and that each person who attempts to take a particular opportunity is equally likely to succeed. Therefore, each citizen's crime rate is \( \sigma(e) = (1/N)[1 - C(m(e))] \).

In Figures 1 and 2 the locus \( \Gamma \) represents the citizens' "reaction function" \( (e, \sigma(e)) \). If crimes are exclusive (\( \mu(m) = 0 \) for all \( m \)), then the crime rate \( \sigma(e) \) is minimized by convicting all defendants, \( e = 0 \). This is the case described in Figure 1. It is only when crime opportunities are nonexclusive that optimal deterrence requires \( e > 0 \). By monotonicity of \( f(e|g)f(e|i) \), \( K(e, p, q) \) is strictly quasi-concave, and there is a unique standard of evidence \( e \) that maximizes

---

4. This is not entirely obvious. In a previous version of this article we assumed that the prosecutor tried only cases with \( e > e_i \) where it chose \( e \) so that a defendant was sent to trial if and only if he would be convicted. If the jury knows this, it updates its information accordingly in such a way that it convicts exactly the same defendants as would be convicted if all cases went to trial.
Figure 1. Citizens' reaction function when crime opportunities are exclusive.

\[ K(e,p,q), \text{ and hence minimizes the crime rate. This standard of evidence, shown in Figure 2, is described in Lemma 1.} \]

**Lemma 1.** If crime opportunities are exclusive, then the crime-minimizing standard of evidence is \( e = 0 \). If crime opportunities are nonexclusive, then the crime-minimizing standard of evidence, which is \( \arg\max K(e,p,q) \), satisfies \( f(e|g)/f(e|i) = p/q \).

We now turn to the jury's problem. As explained in the introduction, we assume that the jury pursues ex post justice. It must decide how to balance the possibility of convicting the innocent (a Type I error) against the possibility of acquitting the guilty (a Type II error). We model this in the now customary way (Rubinfeld and Sappington, 1987), by assuming the jury's objective function is

\[
\text{minimize}_{e \geq 0} \left[ 1 - F(e|i)\ell(i|h)L^1 + F(e|g)\ell(g|h)L^\uparrow \right]
\]

(2)

Figure 2. Citizens' reaction function when crime opportunities are nonexclusive.
where $\ell(i|I)$ and $\ell(g|I)$ represent the jury's prior probabilities of innocence and guilt (before seeing the evidence $e$) conditional on the defendant having been indicted (denoted $I$). $L^1$ and $L^u$ are, respectively, the costs of Type I and Type II errors.

The ratio $\ell(i|I)/\ell(g|I)$ is the jury's prior likelihood ratio of innocence versus guilt, and reflects the probative value of indictment. Of course, jurors are instructed to behave as if the indictment contains no information about guilt or innocence. If the jurors followed these instructions, they would "believe" prior to seeing any other evidence that the accused is no more likely to be guilty than a person pulled randomly off the street. Letting $\text{prob}(g)$ and $\text{prob}(i)$ represent respectively the probabilities that a randomly chosen person is guilty or innocent, the likelihood ratio $\ell(i|I)/\ell(g|I)$ would equal $\text{prob}(i)/\text{prob}(g) = [(N - 1)/N]/(1/N) = (N - 1)$, where $N$ is the size of the population. With this likelihood ratio, the jury would convict only on very strong evidence. We suspect that the jury cannot follow these instructions and, in fact, it would probably not be in society's interest for them to do so.

What then does the ratio $\ell(i|I)/\ell(g|I)$ depend on? We maintain the hypothesis that juries are rational and incorporate their knowledge of the indictment process into their beliefs. Bayesian updating is a strong assumption, but there is no obvious alternative. If jurors do not form their beliefs rationally, how do they form beliefs? The assumption of rationality allows us to see whether one can justify a restriction on jurors' information without appealing to irrationality. If jurors were irrational, one could presumably justify a wide array of restrictions, but it would be unclear what to make of such conclusions.

According to Bayes' Rule,

$$\frac{\ell(i|I)}{\ell(g|I)} = \frac{\text{prob}(I|i) \text{prob}(i)}{\text{prob}(I|g) \text{prob}(g)} = \frac{\ell(I \cap i)}{\ell(I \cap g)}.$$ 

The probability of being indicted conditional on being guilty, $\text{prob}(I|g)$, is $q$, and the probability of being guilty, $\text{prob}(g)$, is $1/N$. To calculate the probability $\text{prob}(I|i)$, we shall think of the random innocent citizen as coming either from the population whose opportunity set includes the crime or from the complement population of innocent citizens. If $n$ people other than the perpetrator have opportunity for the crime, the probability that the innocent citizen had the crime in his or her opportunity set is $n/(N - 1)$. The probability that the citizen drawn randomly from among the innocent did not have opportunity for the crime is $(N - 1 - n)/(N - 1)$. If $\text{prob}(i)/\text{prob}(g) = (N - 1)$, as argued above, then

$$\frac{\ell(i|I)}{\ell(g|I)} = A^p = \frac{pn + p'(N - 1 - n)}{q},$$

where $A^p$ represents the jury's prior likelihood ratio of innocence to guilt.

The jury's best-reply standard of evidence, say $e'(\sigma)$, solves (2) and satisfies
\[ \frac{f(e'|g)}{f(e'|i)} = \Lambda^p \frac{L_I}{L_{II}} \quad \text{as} \quad 0 = e' \quad 0 < e' < \infty. \]

**Proposition 1.** Suppose all citizens have the same opportunity cost of crime, and that \( L^I \geq L_{II} \). Then for all \( \sigma \geq 0 \), \( e'(\sigma) \) is no smaller than the standard of evidence, say \( \hat{e} \), that minimizes the crime rate.

**Proof.** If \( e'(\sigma) = \infty \) or if \( n = 0 \) so that \( \hat{e} = 0 \), the result is immediate. Otherwise (3) satisfies the inequality \( \geq \), and, using \( n \geq 1 \), it follows that \( f(e'(\sigma)|g)/f(e'(\sigma)|i) \geq p/q \). From Lemma 1, \( p/q = f(\hat{e}|g)/f(\hat{e}|i) \), and from monotonicity of the likelihood ratio, it follows that \( e'(\sigma) \geq \hat{e} \). \(\square\)

To investigate the consequences of restricting character evidence we must have criminals of different character. We will therefore repeat this line of reasoning for the case of two subpopulations with different tastes for crime. Subpopulations with different crime rates introduce complexities for the citizens' and jury's best replies. The incentive for a citizen of one type to commit a crime depends on the strategy of the other type, since this affects \( u(m) \) for each \( m \). The incentive for the jury to convict depends on its prior likelihood of guilt versus innocence, which depends on the crime rate of the defendant and also on the crime rates of the other citizens.

The citizen's problem (1) must be considered separately for each type. We will assume that the two types, high crime and low crime (H and L), have different opportunity costs of crime, with \( M_H < M_L \). The payoff to H is as follows, and symmetrically for L:

\[ \pi_H(m,e_H,\mu) = m - M_H - K(e_H,\mu,p,q) \]

Given \((e_L,e_H)\), we will characterize the citizens' best replies by cutoff values \( m_H(e_L,e_H) \) and \( m_L(e_L,e_H) \), which specify the minimum value crime that each type of citizen would commit. As before, if crime opportunities are exclusive, then H (and similarly L) will take a crime opportunity if and only if \( \pi_H(m,e_H,0) \geq 0 \). The difficult case is when crimes may be nonexclusive. Suppose that \( m_H(e_L,e_H) < m_L(e_L,e_H) \). The best replies imply that \( \mu(m) = 0 \) for \( m < m_H(e_L,e_H) \), since no one wants to commit those crimes, and \( \mu(m) = 1 \) for \( m < m_L(e_L,e_H) \) since everyone wants to commit those crimes. For crimes \( m \) such that \( m_H(e_L,e_H) < m < m_L(e_L,e_H) \), \( \mu(m) = \mu_H \), where \( \mu_H \) is the probability that a randomly drawn group of size \( n \) with opportunity for the crime contains at least one type-H person, who will want to commit the crime. Thus best replies \( m_H(e_L,e_H) \) and \( m_L(e_L,e_H) \) are described by:

If crimes are exclusive, \( \pi_H(m_H(e_L,e_H),e_H,0) = 0 \) and 
\[ \pi_L(m_L(e_L,e_H),e_L,0) = 0. \]
If crimes are nonexclusive, and if \( m_H(e_L,e_H) < m_L(e_L,e_H) \),
then \( \pi_L(m_L(e_L,e_H), e_L, 1) = 0 \) and \( \pi_H(m_H(e_L,e_H), e_H, \mu_H) = 0 \).

If crimes are nonexclusive, and if \( m_H(e_L,e_H) > m_L(e_L,e_H) \),
then \( \pi_H(m_H(e_L,e_H), e_H, 1) = 0 \) and \( \pi_L(m_L(e_L,e_H), e_L, \mu_L) = 0 \). (5)

If \( m_H(e_L,e_H) < m_L(e_L,e_H) \), the corresponding best-reply crime rates \( \sigma_L \) and \( \sigma_H \)
are given in (6), along with the total crime rate \( \sigma_T \). If \( m_H(e_L,e_H) > m_L(e_L,e_H) \),
the definitions are reversed in the obvious way. According to (6), \( \sigma_H(e_L,e_H) > \sigma_L(e_L,e_H) \) if and only if \( m_H < m_L \).

\[
\begin{align*}
\sigma_L(e_L,e_H) &= \frac{1}{N_L + N_H} \left( 1 - C(m_L(e_L,e_H)) \right) \\
\sigma_H(e_L,e_H) &= \frac{1}{N_L + N_H} \left( 1 - C(m_L(e_L,e_H)) + \nu_H[C(m_L(e_L,e_H)) - C(m_H(e_L,e_H))] \right) \\
\sigma_T(e_L,e_H) &= 1 - C(m_L(e_L,e_H)) + \mu_H[C(m_L(e_L,e_H)) - C(m_H(e_L,e_H))] \\
\end{align*}
\] (6)

where \( \nu_H < 1 \) is the probability that a given type-H citizen takes a given crime opportunity with \( m \) between \( m_H \) and \( m_L \). This probability reflects the probability that a random citizen has opportunity for that crime, and also the expected number of other type-H citizens with opportunity for that crime, all of whom compete to do it. The variable \( \mu_H \) is the probability that at least one person in a random group of size \( n + 1 \) is type H. In the following lemma, \( \mu_H \) is the probability that at least one person among \( n \) is type H, and therefore it is slightly smaller than \( \mu'_H \).

**Lemma 2.** Suppose crimes are nonexclusive. Let \((e_L,e_H)\) minimize the overall crime rate \( \sigma_T(e_L,e_H) \). Then \( m_H(e_L,e_H) < m_L(e_L,e_H) \), \( \mu_H p/q = f(e_H|g)/f(e_H|i) \), and \( p/q = f(e_L|g)/f(e_L|i) \).

**Proof.** We shall show the two equalities assuming that \( m_H(e_L,e_H) < m_L(e_L,e_H) \). We then observe that exactly the same argument works if \( m_L(e_L,e_H) < m_H(e_L,e_H) \), but we must reverse L and H in the equalities. We then show that in fact \( m_H(e_L,e_H) < m_L(e_L,e_H) \).

A marginal change in \( e_L \) changes only the cutoff \( m_L(e_L,e_H) \), but not \( m_H(e_L,e_H) \). The crime rate can be reduced by increasing \( m_L(e_L,e_H) \), and since \( m_L(e_L,e_H) = M_L + K(e_L,p,q) \), this implies that \( e_L = \arg\max K(e_L,p,q) \). Thus \( p/q = f(e_L|g)/f(e_L|i) \). Similarly, a marginal change in \( e_H \) changes only the cutoff \( m_H(e_L,e_H) \), and not \( m_L(e_L,e_H) \), and since \( m_H(e_L,e_H) = M_H + K(e_H,\mu_H,p,q) \), this implies that \( e_H = \arg\max K(e_H,\mu_H,p,q) \). Thus \( \mu_H p/q = f(e_H|g)/f(e_H|i) \). If \( m_L(e_L,e_H) < m_H(e_L,e_H) \), the same arguments hold with H and L reversed.

Now we show that \( m_H(e_L,e_H) < m_L(e_L,e_H) \). If \( m_L(e_L,e_H) < m_H(e_L,e_H) \),
then \( \pi_H(m,e_H,\mu_H) < \pi_L(m,e_L,\mu_L) \) for \( m \) such that \( m_L(e_L,e_H) < m < m_H(e_L,e_H) \). Since \( M_H < M_L \), this would imply that \( K(e_H,\mu_H,p,q) < K(e_H,\mu_L,p,q) \). But this is a contradiction, since \( e_L \) maximizes \( K(e,\mu_L,p,q) \). 

Lemma 2 characterizes the crime-minimizing \((e_L,e_H)\). But since the jury cares about ex post justice rather than ex ante deterrence, these are not the \((e_L,e_H)\) it will typically choose. We now discuss the jury's best reply to the crime rates \((\sigma_L,\sigma_H)\), which we will designate \(e_H(\sigma_L,\sigma_H)\) and \(e_L(\sigma_L,\sigma_H)\). We must first modify the jury's likelihood ratio to account for the fact that citizens of different types have different crime rates.

The prior probabilities \(\text{prob}(g)\) and \(\text{prob}(i)\) will depend on the defendant's type. If the randomly drawn citizen is type H, \(\text{prob}(g) = \sigma_H/[\sigma_H N_H + \sigma_L N_L]\) and \(\text{prob}(i) = [\sigma_H (N_H - 1) + \sigma_L N_L]/[\sigma_H N_H + \sigma_L N_L]\). Hence \(\text{prob}(i)/\text{prob}(g)\) for a type-H citizen is \([\sigma_H (N_H - 1) + \sigma_L N_L]/\sigma_H\), and symmetrically for a type-L citizen. For convenience we will express the likelihood ratios \(l^L(i)/l^L(g)\) and \(l^H(i)/l^H(g)\) as functions \(A^L\) and \(A^H\) of \(\sigma_L/\sigma_H\). The relative likelihood of innocence and guilt if the defendant is type H is \(A^H\) and it increases with \(\sigma_L/\sigma_H\). The relative likelihood for type L is \(A^L\), which decreases with \(\sigma_L/\sigma_H\).

\[
A^H(\sigma_L/\sigma_H) = \frac{pn + p'(N - 1 - n)}{q} \left( \frac{N_H - 1 + N_L(\sigma_L/\sigma_H)}{N - 1} \right)
\]

\[
A^L(\sigma_L/\sigma_H) = \frac{pn + p'(N - 1 - n)}{q} \left( \frac{N_L - 1 + N_H(\sigma_H/\sigma_L)}{N - 1} \right)
\]

We notice that \(A^P = \gamma^L A^L(\sigma_L/\sigma_H) + \gamma^H A^H(\sigma_L/\sigma_H)\), where \(\gamma_\theta = \sigma_\theta N_\theta/[\sigma_L N_L + \sigma_H N_H]\), \(\theta = L,H\).

By the monotone likelihood ratio property, the jury's objective function is strictly quasi-convex in \(e\). Minimizing (2), where we have substituted \(l^L(i)/l^L(g)\) and \(l^H(i)/l^H(g)\) as functions \(A^L\) and \(A^H\) of \(\sigma_L/\sigma_H\), the jury's best replies \(e_H(\sigma_L/\sigma_H)\) and \(e_L(\sigma_L/\sigma_H)\) satisfy the following conditions:

\[
\frac{f(e_H|g)}{f(e_H|i)} > A^H(\sigma_L/\sigma_H) \frac{L^1}{L^0} \quad \text{as} \quad 0 = e_H < e_H = \infty ,
\]

\[
\frac{f(e_L|g)}{f(e_L|i)} > A^L(\sigma_L/\sigma_H) \frac{L^1}{L^0} \quad \text{as} \quad 0 = e_L < e_L = \infty .
\]

In Lemma 3 and Proposition 2 we use the hypothesis that \((1/n) \leq (N_\theta - 1)(N - 1)\) for \(\theta \in \{L,H\}\). Intuitively this means that a group of \(n\) citizens drawn randomly from a population of size \(N - 1\) will contain on average at least one type-\(\theta\) citizen. Hence if only type-\(\theta\) citizens commit crime in equilibrium, the crime opportunity is still on average "nonexclusive." When a type-\(\theta\) citizen contemplates the crime, he or she knows that on average at least one other type-\(\theta\) has the crime opportunity. We will use this hypothesis in the following sections as well, but when \(n\) must be large for other reasons, we include this assumption in the hypothesis that \(n\) is "sufficiently large."
Lemma 3 is used in Proposition 2, which follows, and in the next section where we compare deterrence in different equilibria, which have different standards of evidence for citizens of different types and according to whether character evidence is admissible. Argmax $K(e,p,q)$ satisfies $f(e|g)/f(e|i) = p/q$, and $K(e,p,q)$ is strictly quasi-concave by the monotone likelihood ratio property. Hence if $e > e' > \text{argmax } K(e,p,q) \geq \text{argmax } K(e,\mu p,q), 0 \leq \mu \leq 1$, then $K(e,p,q) < K(e',p,q) < \text{max } K(e,p,q)$; $e$ is a weaker deterrent than $e'$, which is a weaker deterrent than $\text{argmax } K(e,p,q)$. Since Lemma 3 tells us that the jury’s best replies, and hence the equilibrium standards of evidence, exceed $\text{argmax } K(e,p,q)$, we can compare deterrence in different equilibria by ranking standards of evidence.

Lemma 3. Suppose crime opportunities are nonexclusive, $L^1 \geq L^H$, and $(1/n) \leq (N_0 - 1)/(N - 1)$ for $\theta = L,H$. Let the functions $e_{H}\bullet$ and $e_{L}\bullet$ be the jury’s best-reply functions that satisfy (7). For all $(\sigma_{L},\sigma_{H}),e_{\theta}(\sigma_{L},\sigma_{H}) \geq \text{argmax } K(e,p,q), \theta = L,H.$ Thus

$$\frac{f(e_{\theta}(\sigma_{L},\sigma_{H})|g)}{f(e_{\theta}(\sigma_{L},\sigma_{H})|i)} \geq \frac{p}{q} \quad \text{if } e_{\theta}(\sigma_{L},\sigma_{H}) < \infty.$$  

Proof. We will argue for type H. If $e_{H}(\sigma_{L},\sigma_{H}) < \infty$, the relevant inequality in (7) is $\geq$, which together with the hypotheses implies $f(e_{H}|g)/f(e_{H}|i) \geq \lambda_{H}(\sigma_{L}/\sigma_{H})(L^1/L^H) > p/q$. 

Proposition 2. Let $(\hat{e}_{L},\hat{e}_{H})$ represent the standards of evidence that minimize crime. Suppose $L^1 \geq L^H$ and $(1/n) \leq (N_0 - 1)/(N - 1)$ for $\theta \in \{L,H\}$. For all $(\sigma_{L},\sigma_{H}), \hat{e}_{H} \leq e_{H}(\sigma_{L},\sigma_{H})$ and $\hat{e}_{L} \leq e_{L}(\sigma_{L},\sigma_{H})$.

Proof. The conclusion is immediate if crime opportunities are exclusive, since $\hat{e}_{L} = \hat{e}_{H} = 0$. For the case of nonexclusive crime opportunities, we will give the argument for type H. The conclusion is immediate if $e_{H}(\sigma_{L},\sigma_{H}) = \infty$. For the remaining case, Lemmas 2 and 3 imply that $f(e_{H}(\sigma_{L},\sigma_{H})|g)/f(e_{H}(\sigma_{L},\sigma_{H})|i) \geq p/q \geq f(\hat{e}_{H}|g)/f(\hat{e}_{H}|i).$ The conclusion follows by monotonicity of $f(e|g)/f(e|i)$.

Propositions 1 and 2 imply that the jury will be too lenient relative to maximum deterrence. We did not have to characterize equilibrium in order to prove this, because the jury’s best reply to any crime rates—not just to the equilibrium crime rates—is to set standards of evidence that are more lenient than those that minimize crime. However in order to compare crime rates when character evidence is admissible to crime rates when it is inadmissible, we must characterize both equilibria. We do this in the next section.

3. Character Evidence and Crime Rates
As a benchmark, we begin by supposing that there is no prejudice in the criminal justice system. Thus, the jury’s valuation of the cost of Type I and Type II errors is the same for all citizens and these costs do not depend on the
crime rate—that is, the jury’s loss from erroneously convicting an innocent defendant does not depend on whether or not the defendant is guilty of other crimes. Character evidence at trial is only probative. Furthermore, the police are not prejudiced and therefore all citizens face the same probabilities of indictment for crimes, conditional on being innocent or guilty. In order to assume that citizens’ different incentives for crime originate in exogenous factors rather than in how the criminal justice system treats them, we assume again that $M_H < M_L$.

Our goal is to compare the Nash equilibrium that arises when the jury cannot observe character evidence with the equilibrium that arises when character evidence is admissible. When character evidence is not admissible, the jury chooses a single standard of evidence, $e_P$ (where $P$ stands for 'pooled'), that applies to both types. A Nash equilibrium without character evidence is a vector $(m'_L, m'_H, e'_P)$ that satisfies the best-reply functions (5) and (3). When character evidence is admissible, the jury chooses two standards of evidence, $e_H$ and $e_L$. A Nash equilibrium is a vector $(m^*_L, m^*_H, e^*_L, e^*_H)$ that satisfies the best reply functions (5) and (7). Using (6), we can compute the corresponding equilibrium crime rates.

**Proposition 3.** Let $(m'_L, m'_H, e'_P)$ be a Nash equilibrium when character evidence is admissible, and $(m^*_L, m^*_H, e^*_L, e^*_H)$ be a Nash equilibrium when character evidence is inadmissible. For $n$ sufficiently large, (i) $e^*_L > e'_L > e'_H$; (ii) $m'_L > m^*_L > m^*_H > m'_H$; (iii) $\sigma_L(e'_P, e'_P) < \sigma_L(e^*_L, e^*_H) < \sigma_H(e^*_L, e^*_H) < \sigma_H(e'_P, e'_P)$.

**Proof.** First we show that $e^*_L > e'_L$. Suppose instead that $e^*_L = e'_L$. This implies by (7) that $A^L(\sigma_L(e^*_L, e^*_H)/\sigma_H(e^*_L, e^*_H)) = A^H(\sigma_L(e'_L, e'_H)/\sigma_H(e'_L, e'_H))$, which implies that $\sigma_L(e^*_L, e^*_H) = \sigma_H(e'_L, e'_H)$ and hence $m^*_H \geq m^*_L$. If $m^*_H = m^*_L$, $\pi_H(m^*_H, e^*_H, 1) = \pi_L(m^*_H, e^*_H, 1) = 0$. Using $e^*_L = e'_L$ and Lemma 3, $K(e^*_L, p, q) \leq K(e'_L, p, q)$, which contradicts $\pi_H(m^*_H, e^*_H, 1) = \pi_L(m^*_H, e^*_H, 1) = 0$ since $M_H > M_L$. Similarly, if $m^*_H > m^*_L$, there exists $m$, $m^*_H > m > m^*_L$, such that $\pi_H(m, e^*_L, u_L) < \pi_L(m, e^*_L, u_L)$. Using $e^*_L = e'_L$ and Lemma 3, $K(e^*_L, u_L, p, q) \leq K(e'_L, u_L, p, q)$, so this is also a contradiction. Thus $e^*_L > e'_L$ and from (7), $\sigma_L(e^*_L, e^*_H) < \sigma_H(e'_L, e'_H)$, and $m^*_H < m^*_L$. Since $A^L(\sigma_L/\sigma_H) < A^P < A^L(\sigma_L/\sigma_H)$ when $\sigma_L < \sigma_H$, it also follows from (3) and (7) that $e^*_L > e'_L > e'_H$. Conditions (ii) and (iii) then follow because $K(e'_P, u_H, p, q) < K(e^*_H, u_H, p, q)$ and $K(e'_P, p, q) > K(e^*_P, p, q)$.

Admitting character evidence leads the jury to lower its standard of evidence for habitual criminals and raise its standard of evidence for citizens who are normally law-abiding. The different standards of evidence faced by high-crime and low-crime defendants result from the probative value of the character evidence. The jury is willing to convict habitual criminals on weaker

---

5. For the sake of brevity we omit proofs of existence of Nash equilibria in this game. These proofs are available from the authors. In case of multiple equilibria, the comparative static results that we derive below are valid for all equilibria of this game, except where noted.
evidence because a randomly selected habitual criminal is more likely to be guilty of the crime in question than a randomly selected person who seldom breaks the law. Therefore, the jury begins with a stronger presumption of guilt when the defendant is type H.

Part (iii) of Proposition 3 says that admitting character evidence pulls the different citizens’ crime rates closer together, while excluding character evidence pushes them apart.

Proposition 4. Suppose that \( C(M_L + \max_{e=0} K(e,p,q)) < 1 \). For \( n \) sufficiently large, there is less total crime when character evidence is admissible than when it is inadmissible.

**Proof.** Using (6), the difference in the total crime rate when character evidence is admissible and when not is 
\[
(1 - u_H)[C(m_M^L) - C(m_H^L)] - \mu_H[C(m_M^H) - C(m_H^H)].
\]
By Proposition 3, provided \( C(m_M^H) < 1 \), \( C(m_H^H) - C(m_H^H) \) is positive. If \( C(m_M^H) < 1 \), then \( C(m_H^H) < C(M_H + \max_{e=0} K(e,p,q)) < 1 \). The values of \( m_L^H, m_H^H, m_L^*, m_H^*, m_L^0, m_H^0 \) vary with \( n \), but provided \( C(m_M^H) - C(m_H^H) > \delta \) for some \( \delta > 0 \), then when \( \mu_H \) is close to unity, the difference in total crime rates is negative. The result then follows because \( \mu_H \) increases with \( n \) and equals unity for \( n > N_L \). Thus we only need to argue that \( m_H^0 \) is strictly larger than \( m_L^H \) when \( \mu_H^0 = 1 \). Suppose that as \( n \) becomes large, \( m_L^H \) becomes close to \( m_H^0 \). Then \( e_H^* \) becomes close to \( e_L^p \) and, according to (3) and (7), \( \Lambda_H \) becomes close to \( \Lambda \). It follows from
\[
C(M_L + \max_{e=0} K(e,p,q)) < 1
\]
that \( \sigma_L \) and \( \sigma_H \) are bounded above zero, and therefore the weights \( \gamma_L \) and \( \gamma_H \) are bounded away from zero. Thus \( \Lambda_H, \Lambda^p \) and \( \Lambda^L \) all become close to each other. But then, by (7), \( e_L^* \) also becomes close to \( e_L^* \). If \( e_H^* \) is close to \( e_L^* \), \( m_H^0 \) is bounded below \( m_L^* \), since \( M_H < M_L \). But this means that \( \sigma_H(e_L^*, e_H^*) \) is bounded above \( \sigma_L(e_L^*, e_H^*) \), which contradicts the fact that \( \Lambda_H \) is close to \( \Lambda \).

Proposition 4 means that when opportunity sets are “sufficiently overlapping,” it is irrelevant that type-L citizens commit more crimes when character evidence is admissible, since most of those crimes would have been committed by type-H citizens anyway. Total crime decreases when character evidence is admitted because type-H citizens give up crime opportunities that will not be taken by anyone else. Admitting character evidence reduces crime by decreasing the jury’s leniency toward habitual criminals.

### 3.1 Prejudiced Police

We now suppose that police are prejudiced. Unlike in the previous section, we now assume that citizens are intrinsically identical (i.e., that \( M_H = M_L = M \)) but in equilibrium they face different probabilities of arrest. Thus we will have subscripts H and L on \( p, p', \) and \( q \). Whereas before high-crime and low-crime citizens were so labeled because they were exogenously different in their attitudes toward crime, here we label them high-crime and low-crime because there will be a self-reinforcing equilibrium in which identical citizens are
divided into two crime classes, and those with a higher probability of erroneous arrest will commit more crime in equilibrium.

The jury’s prior likelihood ratios of innocence versus guilt reflect the different probabilities of erroneous arrest. Equation (8) displays the individual likelihood ratios when character evidence is admissible and also the likelihood ratio when the types are pooled.

\[
\begin{align*}
A_H^{\#}(\sigma_L/\sigma_H) &= \frac{p_H n + p_H'(N - 1 - n)}{q_H} N_H - 1 + N_L(\sigma_L/\sigma_H) \\
A_L^{\#}(\sigma_L/\sigma_H) &= \frac{p_L n + p_L'(N - 1 - n)}{q_L} N_L - 1 + N_H(\sigma_H/\sigma_L) \\
A_P^{\#}(\sigma_L/\sigma_H) &= \frac{c_L}{c_L + c_H} A_H^{\#}(\sigma_L/\sigma_H) + \frac{c_H}{c_L + c_H} A_L^{\#}(\sigma_L/\sigma_H),
\end{align*}
\]  
(8)

where \(c_\theta = \sigma_\theta q_\theta N_\theta, \theta \in \{H, L\} \).

We consider two cases: first, that \(p_H' > p_L'\); second, that \(p_H > p_L\), \(q_H < q_L\). In the first case, the two types' different incentives for crime arise entirely through the fact that the jury will present them with different standards of evidence. While \(p_H'\) and \(p_L'\) do not affect the citizens' decisions directly, they do affect the equilibrium standard of evidence. But in the second case, a larger ratio \(p/q\) increases incentives for crime even for a fixed standard of evidence. We want to compare crime rates in populations where the police "bring in the usual suspects" with crime rates where the police screen more carefully. Randomness in the arrest of type-H citizens is captured in the hypothesis that \(p_H/q_H > p_L/q_L\). But if the police use the same enforcement budget per capita in the two populations, it must also be the case that \(p_H > p_L\) and \(q_H < q_L\). These two circumstances imply that \(K(e, q_H, p_H) \leq K(e, q_L, p_L)\) for all \(e\).

In both cases, when the citizens have the same crime rates, the jury believes that a type-H defendant is more likely to be innocent than a type-L defendant: \(A_H^{\#}(1) > A_L^{\#}(1) > A_P^{\#}(1)\).

As before, we compare the equilibrium \((m_L^*, m_H^*, e_P^*)\) that arises when the jury cannot observe character evidence with the equilibrium \((m_L^*, m_H^*, e_P^*, e_H^*)\) that arises when character evidence is admissible. Once again these equilibria are defined using the best replies given in (5) and (7), but the jury’s likelihood ratios are now given by (8). In the following proposition we compare the equilibria when character evidence is admissible and when not.

**Proposition 5.** (a) Suppose that \(p_H' > p_L'\), and that \(p_H = p_L\), \(q_H = q_L\). Then for \(n\) sufficiently large, conditions (i), (ii), and (iii) hold.

(i) \(e_P^* < e_P < e_H^*\),
(ii) \(m_L^* > m_L = m_H > m_H^*\),
(iii) \(\sigma_H(e_L^*, e_H^*) > \sigma_L(e_P, e_P) = \sigma_H(e_P, e_P) > \sigma_L(e_L^*, e_H^*)\).

(b) Suppose that \(p_H > p_L\) and \(q_H < q_L\). Then if \(p_L\) and \(p_L'\) are sufficiently close to zero, if \(n\) is sufficiently large, and if \(C(M + \max_{e=0} K(e, p_L, q_L)) < 1\), conditions (i), (iv) and (v) hold.
Proof. (a) First consider an equilibrium with character evidence. If \( e^*_H \leq e^*_L \),
then (7) implies \( \Lambda^*_H \leq \Lambda^*_L \), which, using (8), implies \( \sigma_H(e^*_L,e^*_H) > \sigma_L(e^*_L,e^*_H) \).
Using Lemma 3, \( e^*_L \geq e^*_H \geq \arg \max K(e,p,q) \), and therefore \( m^*_L \leq m^*_H \), which
contradicts the implication that \( \sigma_H(e^*_L,e^*_H) > \sigma_L(e^*_L,e^*_H) \). Thus \( e^*_L < e^*_H \), \( m^*_L \leq m^*_H \), and \( \sigma_H(e^*_L,e^*_H) > \sigma_L(e^*_L,e^*_H) \). Now consider the equilibrium without character evidence. Since \( p_\theta = \theta = L,H \), does not enter the individual’s decision problem, \( \sigma_L(e^*_L,e^*_H) = \sigma_L(e^*_L,e^*_H) \).
By (7) and (8), \( e^*_L < e^*_H \), which (using Lemma 3) implies (ii) and (iii).

(b) Lemma 3 must be modified slightly when \( p/q \) differs for the two types.
When \( n \) is sufficiently large, \( f(e^*_H) / f(e^*_L), f(e^*_L) / f(e^*_H), f(e^*_L) / f(e^*_H) = p_H / q_H \), and \( e^*_L, e^*_H, e'_L \) arg \( \max K(e,p_H,q_H) \), therefore \( K(e,p_H,q_H) \) and \( K(e,p_H,q_L) \) are decreasing in \( e \) for \( e \geq \min \{ e^*_H, e^*_L, e'_H \} \).

The hypothesis on \( M \) implies that \( \sigma_L \) and \( \sigma_H \) are bounded above zero. If \( p_L \) and \( p'_L \) are small, we can assume that \( \Lambda^*_H(e^*_L,e^*_H) / \sigma_H(e^*_L,e^*_H) \).
Therefore, according to (7), \( e^*_L < e^*_H \), and since \( K(e^*_L,p_H,q_H) < K(e^*_L,p_L,q_L) \), it follows that \( m^*_L > m^*_H \) and \( \sigma_L(e^*_L,e^*_H) < \sigma_H(e^*_L,e^*_H) \). According to the last line of (8), \( \Lambda^*_L(e^*_L,e^*_H) / \sigma_L(e^*_L,e^*_H) > \Lambda^*_H(e^*_L,e^*_H) / \sigma_H(e^*_L,e^*_H) \), and therefore, according to (7), \( e^*_L < e'_L < e^*_H \). Hence \( m^*_L > m^*_L > m^*_H > m^*_H \) and \( \sigma_H(e^*_L,e^*_H) > \sigma_H(e^*_L,e^*_H) > \sigma_L(e^*_L,e^*_H) > \sigma_L(e^*_L,e^*_H) \).

Character evidence leads the jury to use a higher standard of evidence for habitual criminals and a lower one for normally law-abiding citizens. This is because the jury knows that a type-H defendant is more likely to be arrested when innocent, and therefore it is more likely that he is innocent at trial. The higher standard of evidence weakens deterrence. The police’s willingness to arrest high-crime citizens presumably reflects a desire to deter crime, but instead has the opposite effect. Because of the jury’s rational response, the high-crime citizens are reinforced in their life of crime and have a higher crime rate than if character evidence were inadmissible at trial.

It is not immediately clear from Proposition 5 whether excluding character evidence increases or decreases total crime, but the following proposition states that when crime opportunities are sufficiently overlapping, crime is reduced by disallowing character evidence.

Proposition 6. Suppose either that \( p'_H > p'_L \), and \( p_H = p_L, q_H = q_L \) or that \( p_H > p_L \) and \( q_H < q_L \). Suppose also that \( C(M + \max_{e=0} K(e, p_L, q_L)) \) < 1. For \( n \) sufficiently large, excluding character evidence reduces crime.

Proof. As in the proof of Proposition 4, it is enough to show that \( C(m^*_H - C(m^*_L)) < \delta \) for some \( \delta < 0 \), when \( \mu^*_H \) is close to one. Suppose that as \( n \) becomes large, \( m^*_H \) becomes close to \( m^*_H \). Then \( e^*_L \) becomes close to \( e'_L \) and, according to (7), \( \Lambda^*_H \) becomes close to \( \Lambda^*_H \). For each \( \sigma / \sigma_H, \Lambda^*_H \) is a convex
combination of $A_L^*$ and $A_H^*$ with coefficients bounded away from zero, and therefore $A_H^*$ becomes close to $A_L^*$. But then, by (7), $e_H^*$ also becomes close to $e_L^*$. If $e_H^*$ is close to $e_L^*$, $m_H^*$ is bounded below $m_L^*$, since $K(e_H^*, p_H, q_H)$ is strictly smaller than $K(e_L^*, p_L, q_L)$. But this means that $\sigma_H(e_L^*, e_H^*)$ is bounded above $\sigma_L(e_L^*, e_H^*)$, which implies that $A_H^*$ is bounded above $A_L^*$, a contradiction.

When crime opportunities are sufficiently overlapping, admitting character evidence increases crime by reinforcing the jury’s leniency toward habitual criminals, who face a higher probability of erroneous arrest. This leniency increases these high-crime citizens’ incentive to commit crime. Their crime rates are increased further by their willingness to step in and take most of the crime opportunities that the rest of society—which faces a lower standard of evidence—now forgoes. Thus, total crime increases.

3.2 Prejudiced Jury

We now suppose that the jury is prejudiced against habitual criminals. We take this prejudice to mean that the jury’s cost of Type I errors, $L^1$, depends on the defendant’s crime rate. The jury feels a greater loss if it wrongly convicts a noncriminal than if it wrongly convicts a habitual criminal. Thus, we assume that $L^1$ depends on $\sigma$, that $L^1(\sigma)$ is decreasing, that $L^1(0) = \infty$, and that there exists $\sigma$ such that $L^1(\sigma) = 0$ for $\sigma > \sigma$. It is no longer true that the jury always chooses a standard of evidence $e$ that is greater than the deterrence-maximizing standard, since the hypothesis that $L^1 \geq L^0$ is not necessarily satisfied. In fact, the reverse problem may arise. If citizens have overlapping crime opportunities, the jury may choose a standard of evidence that is below the deterrence-maximizing level.

Figures 3, 4, and 5 show the jury’s “reaction functions” $J_L$, $J_H$, and $J_p$ that satisfy (7), where the likelihood ratio depends on whether character evidence is admissible (and on the defendant’s type) and $L^1$ depends on $\sigma$. If the jury is convinced that the defendant never commits crime in equilibrium (i.e., $\sigma = 0$), then $L^1 = \infty$ and the jury will not convict the defendant for the crime at hand. Hence as $\sigma$ becomes close to zero, the jury’s best-reply standard of evidence becomes large. But if the defendant has a high crime rate, in particular if $\sigma > \sigma$, then the jury will convict the defendant whatever the evidence, and their best-reply standard of evidence is zero.

With jury prejudice it is not always true that restricting character evidence reduces crime, but it does so under the circumstances summarized in Lemmas 4 and 5. Lemma 4 summarizes conditions under which there is a “punitive” equilibrium in which the jury is willing to convict some defendants on the basis of their character alone without consulting the evidence. Excluding character evidence clouds the jury’s information and makes them more lenient toward habitual criminals. This leniency improves deterrence by increasing the expected difference in costs when guilty versus when innocent, $K(e, q, \mu_H P)$.

The punitive equilibrium when character evidence is admissible is summarized in Lemma 4 and drawn in Figures 3 and 4. Figure 3 describes the crime rate of type-L citizens. Since we assume in Lemma 4 that $M_L$ is high, the
citizens’ reaction function $\Gamma_L$ is pushed far to the left. If $M_L$ is high, for each standard of evidence $e$ the best-reply $m_L$ is relatively high and the corresponding crime rate $\sigma_L$ is relatively low. The two assumptions on $M_L$ in Lemma 4 mean that at the crime-minimizing standard of evidence the crime rate of type-$L$ citizens is zero, and at $e = 0$ it is less than $\sigma$. It follows from the shape of the jury’s reaction function that $e^L > \text{argmax} K(e, p, q)$.

Figure 4 describes the equilibrium crime rate of type-$H$ citizens. There may be multiple equilibria, but our argument applies for the equilibrium in which type-$H$ citizens take all their crime opportunities, shown by the large dot in Figure 4, where “maximum crime rate” means the crime rate when $C(m^*_H) = 0$. [The crime rate of type-$H$ depends on $m^*_L$ as well as on $m^*_H$, since the two types “compete” for crimes with $m > m^*_L$. See Equation (6).] Such an equilibrium exists if $M_H$ is sufficiently low and the crime rate of type-$H$ citizens
would be at least \( \sigma \) when \( e_H^p = 0 \). This is ensured by the hypotheses on \( M_H \) and \( \sigma \) in Lemma 4, whatever the crime rate of type L.

**Lemma 4.** Suppose that crime opportunities overlap and character evidence is admissible. Suppose that \( L(L^0(0) = \infty, (1/N_L + N_H)[1 - C(M_L + K(0,p,q))] < \sigma \leq (1/N_L + N_H), C(M_L + \max_{e^p \in \Omega} K(e,p,q)) = 1, \) and \( C(M_H + K(0,u_H p,q)) = 0 \). Then there is an equilibrium \((e_H^L, e_H^p, m_H^L, m_H^p)\) where \( e_H^p = 0, e_H^p > \text{argmax} \, K(e,p,q) \), and \( C(m_H^L) = 0 \).

Now consider the equilibrium \((m_L', m_H', e_L')\) without character evidence. Figure 5 shows the reaction functions of the jury and citizens, as well as the equilibrium standard of evidence. The \( \sigma_p \) on the horizontal axis represents the expected crime rate of a defendant when the jury cannot observe the defendant's type. The jury's reaction function \( J_p \) satisfies (7), where the likelihood ratio is \( \Lambda^p \) and \( L^1 \) decreases with \( \sigma_p \). If \( \sigma_p < \sigma < \sigma_H(e_H^L, e_H^p) \), then \( \Lambda^p < \Lambda(L^1(\sigma_H(e_H^L, e_H^p))) \), and therefore the jury's reaction function \( J_p \) lies below \( J_L \) for all such \( \sigma_p \) in Figure 5. The citizens' pooled reaction function \( \Gamma_p \) is \((e_p, \sigma_p(e_p))\), where \( \sigma_p(e_p) \) is the expected crime rate of defendants at trial: \( \sigma_p(e_p) = \phi(L|L)\sigma_L(e_p, e_p) + \phi(H|L)\sigma_H(e_p, e_p) \), where \( \phi(L|L) \) and \( \phi(H|L) \) are respectively the probabilities that the defendant is type L or H conditional on having been indicted. [The expected crime rate of the defendant at trial is not a simple multiple of the total crime defined in (6) because the probabilities of ending up at trial are not the same for both types.] Since \((1/N_H)\) is an upper bound for the crime rate of type-H citizens, \( \sigma_p(e) \leq \phi(L|L)\sigma_L(e, e) + \phi(H|L)(1/N_H) \). If \( N_L/N_H \) is large, \( \phi(H|L) \) is small, and \( \Gamma_p \) is close enough to \( \Gamma_L \) to ensure that \( e_p > \text{argmax} \, K(e,p,q) \). Thus we have Lemma 5.

![Figure 5. Standards of evidence. In the pooled equilibrium, the jury's standard of evidence is \( e_H^p \), the expected crime rate of a citizen at trial is \( \sigma_p(e_p) \), and type L's crime rate is \( \sigma_L(e_L^*, e_p) \). In the separated equilibrium, the jury's standard of evidence for type L defendants is \( e_L^* \) and the equilibrium crime rate for type L is \( \sigma_L(e_L^*, e_p) \). \( \sigma_L(e_L^*, e_p) > \sigma_L(e_L^*, e_p) \) because \( e_L^* > e_p \).](image-url)
Lemma 5. Let $M_L$ and $N_L/N_H$ be large enough so that if $e < \text{argmax } K(e,p,q)$, then $e_p(\sigma_*(e)) > e$. Let $(m'_L, m'_H, e'_p)$ be an equilibrium without character evidence. Then $e'_p > \text{argmax } K(e,p,q)$.

To show that total crime decreases when character evidence is excluded, it is enough to show that $m'_L < m_L$, which means that type-L citizens take fewer crime opportunities when character evidence is excluded than when it is allowed. Type-H citizens automatically take fewer crime opportunities when character evidence is excluded, since they were taking all possible crime opportunities in the equilibrium with character evidence ($C(m_H^*) = 0$). In Figure 5 the reaction function $\Gamma_p$ lies to the right of $\Gamma_L$, since at each standard of evidence $e_p$ type H has a higher crime rate than type L. Since $e'_p > \text{argmax } K(e,p,q)$, and since $J_L$ is below $J_L$, it follows that $e'_p < e_L^*$. The latter ensures that $m'_L < m_L$, and Proposition 7 follows.

Proposition 7. Let $M_L$ and $N_L/N_H$ satisfy the hypotheses in Lemmas 4 and 5, and let $\sigma$ satisfy the hypothesis in Lemma 4. Let $(e_L^*, e_H^*, m_L^*, m_H^*)$ be an equilibrium with character evidence in which $e_H^* = 0$ and $e_L^* > \text{argmax } K(e,p,q)$. Let $(m'_L, m'_H, e'_p)$ be an equilibrium without character evidence. Then the total crime rate without character evidence is smaller than the total crime rate with character evidence.

Proposition 7 states that under the conditions outlined in Lemmas 4 and 5 excluding character evidence at trial improves deterrence. This happens for two reasons. First, the jury is less lenient toward normally law-abiding citizens, which decreases their incentive to commit crime. Second, the jury will no longer convict habitual criminals automatically, but instead will allow the evidence to discriminate guilt from innocence. For habitual criminals, excluding character evidence increases the expected difference in costs between guilt and innocence, and therefore they find crime less attractive.

4. Conclusion

There is a potential conflict between the goal of ex post justice and the goal of ex ante deterrence, even though both goals are served by convicting the guilty and acquitting the innocent. The potential conflict is that, because the jury cares about ex post errors that have no bearing on deterrence, the jury will typically choose a standard of evidence other than the one that maximizes deterrence. If the jury is not prejudiced, it will typically employ a higher standard of evidence (be more lenient) than the standard of evidence that minimizes crime. This conclusion supports the prevailing public sentiment on crime, which holds that the criminal justice system should be more punitive toward habitual criminals, such as members of gangs.

We conclude our investigation by summarizing the circumstances where excluding character evidence reduces crime and where not. First, when there is neither police prejudice nor judicial prejudice in the hierarchy, and when
crime opportunities overlap substantially, excluding character evidence increases crime. Restrictions on character evidence cause the jury to use the same standard of evidence for all defendants. Citizens with a low opportunity cost of crime are pooled with other citizens, and face a more lenient jury than when character evidence is admissible. Therefore excluding character evidence makes crime more attractive. On the other hand, citizens with a high opportunity cost of crime face a tougher jury than with character evidence, and find crime less attractive. But if crime opportunities are sufficiently overlapping, the reduction in crime by low-crime citizens does not affect the total crime rate very much. On average at least one high-crime citizen has opportunity for each crime, and therefore those citizens determine the total crime rate. Thus by increasing the attractiveness of crime to high-crime citizens, excluding character evidence increases crime.

But the foregoing argument is overturned if the jury is prejudiced. The argument rests on the supposition that a higher standard of evidence (a more lenient jury) increases the incentive for crime. When the jury is not prejudiced, that supposition is correct, since the jury is already too lenient. In contrast a prejudiced jury can be “too punitive,” in the sense of being willing to convict a habitual criminal irrespective of the evidence linking that person to the crime. If that is what the habitual criminal expects, he or she has little incentive to avoid crime. Restricting character evidence makes the jury less punitive and restores deterrence by making the jury consult the evidence.

We also investigated the effect of character evidence when the police screen some types of citizens more carefully than others before they make an arrest. Those who are not screened carefully are in greater jeopardy of being arrested for crimes they did not commit. A rational jury will assign such a defendant a higher probability of innocence conditional on the evidence, and be less willing to convict him. The jury’s leniency reinforces crime, so that the police’s prejudice has an effect opposite to its apparent intent. However if character evidence is excluded, the jury cannot use a higher standard of evidence for citizens subject to police prejudice. With a lower standard of evidence, such citizens will take fewer crime opportunities, and according to the reasoning outlined above, the total crime rate will be lower.

Thus when either the police or the jury are prejudiced, and when crime opportunities overlap, excluding character evidence can reduce crime.

How believable are the hypotheses that underlie our conclusions? The hypothesis of overlapping crime opportunities is probably descriptive of crime and violence in poor urban communities, but not of white collar crime. Citizens of such communities have few attractive lawful opportunities, so $M_H$ is small. Sociological studies of crime, gangs, and inner-city life provide at least anecdotal evidence that residents in poor urban neighborhoods face the possibility of arrest even if they avoid crime, which means that $\mu_{\text{HH}}p$ is close to $q$. The reasons for this are complex and stem in part from the volume of crime, which makes it more likely that a person will be observed at a crime scene (see Kotlowitz, 1991: 266–68). Other factors, such as association with
a gang and common styles of dress, also contribute to the possibility of erroneous arrest (see Anderson, 1990: 195–98). We conclude from this anecdotal evidence that our hypotheses may approximate reality in these communities.

Of course, some may question the applicability of a model of rationality to study urban crime and violence, which often seems random and unrestrained. We contend, however, that an assumption of rationality may not be as heroic as it sounds. Jankowski (1991: 252–83) describes how inner-city gangs pursue their goals in ways that explicitly recognize the constraints established by the criminal justice system. For example, they exploit the differences between the juvenile and adult justice systems by “assigning” crimes to children. Furthermore, gangs pursue their (illegal) economic activities much like any legal business, comparing risks and rewards (Jankowski, 1991: 101–36). Such behavior is the hallmark of rational decision-making.

References


