Criminal procedure

* Thomas J. Miceli*

Introduction
Criminal procedure is the legal process by which a criminal defendant's guilt or innocence is ascertained. It thus encompasses the set of procedures and rules that apply to criminal defendants from the time of arrest through the adjudication of guilt, and, in the event of a conviction, the imposition of a legal sanction. Key features of this process include pretrial bargaining between the defendant and prosecutor (potentially resulting in a plea bargain), the bail system, the standard for determining guilt at trial, rules of evidence (including constitutional protections of the defendant’s rights), and the appeal process. An economic approach to criminal procedure views these various elements within a common framework whose goal is to minimize the sum of the cost of legal error plus the administrative cost of operating the system.1

Regarding legal error, two types are possible: type 1 errors, or conviction of innocent defendants, and type 2 errors, or acquittal of guilty defendants. We will argue that the rules governing conduct of a criminal trial can best be understood as reflecting a strong bias against committing type 1 errors. This bias, for example, explains the strict standard of proof for conviction (beyond a reasonable doubt), as well as the various legal protections afforded to defendants, such as the exclusionary rule and the right against self-incrimination.2 As for administrative costs, the pervasive use of plea bargaining to dispose of criminal cases in the American

* I acknowledge the insightful comments of Richard Adelstein and Nuno Garoupa.

1 For previous economic analyses of criminal procedure that adopt a similar perspective, see Posner (2003, chapter 21), Lewisch (2000), Ehrlich (1998) and Wittman (1974). This chapter focuses on the American procedural system, though the basic elements are common to European (civil law) systems as well. One difference is the greater prevalence of plea bargaining in the United States, as will be discussed.

2 But see Hylton and Khanna (2007), who offer a different rationale for the generally pro-defendant bias in criminal procedure. Specifically, they argue that this bias discourages individuals and the government from using criminal law to pursue their own self-interest, thereby saving on rent-seeking costs.
The criminal justice system can best be understood as an effort to conserve scarce prosecutorial and judicial resources.

Since criminal procedure concerns the adjudication of guilt, it is a key component of the overall process of law enforcement. Thus, an economic approach to procedure must also recognize its impact on deterrence. To some extent, this creates conflicts because the optimally deterring sanction will not always coincide with the sanction that best serves the other goals of the procedural system. The ongoing debate regarding the appropriate scope of judicial discretion in sentencing reflects this conflict (Miceli, 2008; Shepherd, 2007).

The remainder of the chapter is organized as follows. The next section reviews competing perspectives on how the procedural system as a whole operates – specifically, is it more like a market or is it more like a regulatory system? Subsequent sections then turn to specific aspects of the system. We first consider plea bargaining, which is the manner in which the vast majority of criminal convictions are achieved in the American legal system. We then consider the bail system, which sets the terms under which a criminal defendant can remain free. We next turn to the trial process, focusing on the standard of proof and rules of evidence. Finally, we consider the implications of criminal procedure for deterrence.

**Criminal procedure: competing perspectives**

Easterbrook (1983) has offered an economic perspective on criminal procedure that views it like a market system whose goal is to establish the efficient ‘price’ of crime. He argues that the considerable discretion of prosecutors in determining what cases to pursue, and on what charge, as well as the sentencing discretion of judges, are key elements of this system in that they allow these agents to make decentralized decisions based on their case-specific information. The result is a pricing system that operates like a market to achieve efficient deterrence of crime. It follows that efforts to curtail discretion, such as mandatory sentencing guidelines, can only interfere with the smooth operation of the system, in the same way that government regulations impede the efficient operation of otherwise competitive markets.

Schulhofer (1988) has criticized this market-oriented view of criminal procedure by emphasizing the unfairness in sentencing that is a necessary by-product of unlimited prosecutorial and judicial discretion. Specifically,

---

3 For general models of law enforcement that focus on deterrence, see Becker (1968) and Polinsky and Shavell (2000a). For a model that explicitly incorporates some aspects of the procedural system, see Miceli (1991).
he argues that ‘[d]iscretion leads to unwarranted disparities in the treatment of similarly situated offenders’ (Schulhofer, 1988, p. 80). Moreover, discretion actually undermines deterrence, first, by creating ex ante uncertainty among potential offenders regarding the consequences of their actions, and second, by preventing judges from imposing harsh sentences on especially dangerous offenders. As a result, the market ‘fails’ to establish the efficient price for crime.

More fundamentally, Schulhofer challenges the market analogy by arguing that the procedural system lacks the basic equilibrating mechanisms that are crucial to the determination of market prices. In particular, agents do not receive signals of departures from the equilibrium price, nor do they have incentives to act in ways that restore equilibrium. Further, Garoupa and Stephen (2007) point to the existence of significant market failures inherent in the system, such as asymmetric information and agency costs. Given these problems, limits on judicial and prosecutorial discretion may therefore be necessary to promote efficiency in sentencing, suggesting that the procedural system is better viewed as a regulatory system than as a market.

Adelstein (1981) offers an institutional perspective that views the procedural system as pursuing the objective assumed by Easterbrook – namely, setting the efficient price for crime – but recognizes the imperfections identified by Schulhofer. The result is an evolutionary process (rather than a market) that involves an interplay between legislatures, which seek to send clear signals to would-be offenders ex ante, and prosecutors and judges, who seek to fit punishments to crimes ex post. Whereas predictability promotes deterrence, discretion promotes proportionality, both of which are goals of the criminal process, but which, given the imperfections of criminal enforcement, are generally in conflict. Adelstein’s theory therefore suggests that the procedural system continually evolves, albeit imperfectly, to maintain a balance between these goals. Wittman’s (1974) theory of procedure highlights a similar trade-off between efficiency and fairness in the rules determining criminal punishment.

**The economic objective of the procedural system**

It is not essential to accept the argument that the procedural system has evolved according to market forces in order to view it as having an economic logic. This logic, whether it arose spontaneously or by design, involves maximizing an objective function that includes (1) the benefit of punishing the guilty (which may include deterrence and/or retribution), (2) the cost of punishing the innocent, and (3) administrative costs. Unfortunately, the first two of these goals are fundamentally in conflict, given uncertainty about a defendant’s guilt. This uncertainty,
and the problems it poses for the procedural system, can be illustrated as follows.

When a suspect is apprehended, the prosecutor cannot observe whether he is in fact guilty of the crime in question. Instead, he or she observes an imperfect signal, $\theta$, which can be interpreted as the probability that the defendant is guilty based on the available evidence. (Equivalently, $\theta$ can be interpreted as the defendant’s probability of conviction at trial.) We will refer to $\theta$ as the ‘strength of the case’ against the defendant. We assume that, on average, evidence gathering is effective in the sense that the stronger is the case against a defendant, the more likely it is that he is truly guilty. To capture this, we suppose that $\theta$ is a random variable drawn from a distribution $F_j(\theta)$, where $j$ denotes the defendant’s true status: $j = G$ indicates that he is truly guilty and $j = I$ indicates that he is truly innocent. Further, we assume that

$$F_I(\theta) > F_G(\theta) \text{ for all } \theta. \quad (5.1)$$

That is, the distribution is shifted rightward (in the sense of first-order stochastic dominance) for truly guilty defendants. It follows that observing a high value of $\theta$ is more likely if the defendant is truly guilty, but it is not impossible when the defendant is innocent (as when there is strong circumstantial evidence or if the defendant has been framed). Conversely, a low $\theta$ is more likely if the defendant is innocent, but not impossible if he is guilty (as when he has covered his tracks well or has framed someone else).

In this setting, it should be evident that reaching a verdict based solely on $\theta$, whether by plea or trial, will result in two types of errors: conviction of the innocent, referred to as a type 1 error, and acquittal of the guilty, referred to as a type 2 error. For example, if all defendants with $\theta$ exceeding a threshold $\hat{\theta}$ are found guilty, then the probability of a type 1 error is $1 - F_I(\hat{\theta})$, while the probability of a type 2 error is $F_G(\hat{\theta})$. Further, if the only policy lever available is the standard of guilt, $\hat{\theta}$, then it is impossible to lower one of these probabilities without raising the other (see Figure 5.1). In other words, there is a fundamental trade-off between the two types of errors.

With this trade-off in mind, we first consider the primary manner in which guilt is determined in the American criminal justice system: the plea bargain. Later we consider the determination of guilt by trial.

**Plea bargaining**

Early models of plea bargaining viewed it primarily as a means for the prosecutor to conserve scarce resources by avoiding the cost of having to take all defendants to trial (Landes, 1971). In this perspective, the prosecutor and defendant reach a mutually beneficial agreement whereby
the defendant pleads guilty in return for a lesser penalty than he would have received on conviction at trial (the so-called sentence discount). This outcome is the result of a bargaining process in which each side acts to further his or her own self-interest.

Landes assumed that the prosecutor maximizes the expected punishment imposed on the defendant, less the cost of trial. The prosecutor is therefore concerned with achieving the greatest possible punishment, given his resource constraint, regardless of the factual guilt or innocence of the defendant. For a defendant whose probability of conviction at trial is $q$, the prosecutor’s expected return from trial is thus $q s - C_p$, where $s$ is the penalty on conviction, and $C_p$ is the cost of trial. Given this objective, the prosecutor would only be willing to let the defendant plead guilty in return for a sentence of $s_0$ if $s_0 \geq q s - C_p$.

The defendant, in turn, will only accept this offer if it entails a lower cost than going to trial. The expected cost of trial for the defendant is his expected punishment plus legal costs, or $q s + C_d$. He will therefore only accept the offer of $s_0$ if $s_0 \leq q s + C_d$. A mutually acceptable plea offer is

---

4 Grossman and Katz (1983) and Reinganum (1988) adopt an objective function for the prosecutor that also includes the cost of wrongful punishment, thus presumably bringing the objective of the prosecutor more in line with that of society.
therefore possible if there exists an \( s_0 \) such that \( \theta s + C_d \geq s_0 \geq \theta s - C_p \). It is clear that such an \( s_0 \) always exists since \( \theta s + C_d > \theta s - C_p \). Intuitively, because the parties agree on the expected outcome at trial, \( \theta s \), they have a joint interest in avoiding the trial costs. The model therefore predicts that a large percentage of cases should result in a plea bargain, which is in fact what we observe.

The actual sentence emerging from the plea bargain in this model is ambiguous, as it depends on the bargaining strengths of the parties. All we can say with certainty is that it will lie somewhere between the ‘threat points’ \( \theta s - C_p \) and \( \theta s + C_d \). As an example, the Nash bargaining solution, which gives each player half of the surplus, implies a sentence of

\[
s_0 = \frac{\theta s + (C_d - C_p)}{2}.
\]

Note that this may be larger or smaller than \( s \), the penalty on conviction at trial. It will tend to be smaller the lower is the probability of conviction, \( \theta \), the smaller is the defendant’s cost of trial, \( C_d \), and the larger is the prosecutor’s cost of trial, \( C_p \).

A shortcoming of the preceding model is that it does not explain why cases ever go to trial. One explanation is that the plaintiff and defendant disagree about the probability of conviction at a trial. Specifically, let \( \theta_p \) and \( \theta_d \) be the prosecutor’s and defendant’s assessments, respectively, of the probability of conviction. The condition for a plea bargain to be possible now is \( \theta_d s + C_d \geq \theta_p s - C_p \), or

\[
C_p + C_d \geq (\theta_p - \theta_d)s.
\]

Thus, if \( \theta_p - \theta_d \) is positive and large, a plea bargain is no longer feasible. Intuitively, if both sides are optimistic about their chances of winning at trial (meaning that \( \theta_p \) is high and \( \theta_d \) is low), then a mutually acceptable plea agreement will not exist (Cooter and Rubinfeld, 1989).

A different explanation for why some trials occur is asymmetric information (Reinganum, 1988). Suppose, in particular, that defendants know their own probability of conviction at trial, but the prosecutor only observes, say, the average probability across all defendants, \( \bar{\theta} \). The condition for a mutually acceptable plea bargain in this case is

\[
\theta s + C_d > \bar{\theta}s - C_p,
\]

or

\[
\theta > \bar{\theta} - \frac{C_p + C_d}{s}.
\]
Thus, only defendants with relatively high probabilities of conviction will accept a bargain, while those with low probabilities will opt for trial. In this way, plea bargaining achieves a kind of ‘sorting’ of defendants based on the strength of the cases against them. Note that both the ‘mutual optimism’ and asymmetric information models predict that trials are more likely the larger are the stakes of the case ($s$), and the lower are the costs of trial ($C_p + C_d$).

**Plea bargaining as a sorting device**

The possibility of sorting implied by the asymmetric information model suggests that, in addition to saving on trial costs, plea bargaining might also be used to reduce legal error. Based on this insight, Grossman and Katz (1983) developed a model in which plea bargaining achieves perfect sorting of innocent and guilty defendants by means of self-selection. To illustrate, suppose that all guilty defendants have a probability of conviction at trial equal to $\theta_G$, while all innocent defendants have a probability of conviction equal to $\theta_I$, where $\theta_G > \theta_I$. In this setting, any plea offer $s_0$ that satisfies $\theta_G s + C_d > s_0 > \theta_I s + C_p$ will be acceptable to guilty defendants but not acceptable to innocent defendants. As a result, all guilty defendants will plead guilty, while all innocent defendants will go to trial. In other words, a perfect separating equilibrium occurs in which no innocent defendants are induced to plead guilty.5

There are, however, two significant drawbacks of this outcome. First, it implies that any convictions at trial are wrongful since only the innocent opt for trial. One response is simply for the court to acquit all defendants, but this would undermine the equilibrium because guilty defendants would no longer find it preferable to plead guilty. Thus, in order to support the equilibrium, the court must knowingly convict some innocent defendants (that is, there must be some type 1 errors), or sorting must be imperfect (that is, some guilty defendants must opt for trial).

The second problem with the perfect sorting equilibrium is that, in the real world, innocent and guilty defendants cannot be neatly partitioned based on their probabilities of conviction. This is readily apparent from the specification depicted in Figure 5.1 where the distributions of $\theta$ for innocent and guilty defendants overlap. As a result, it is not possible to choose a plea offer $s_0$ that can perfectly separate those who are factually innocent

---

5 The possibility that some innocent defendants might plead guilty is one of the concerns courts have voiced when considering the acceptability of plea bargaining. See, for example, *Brady v. United States*, 397 US 742 (1970). Also see Adelstein (1978) and Garoupa and Stephen (2007).
from those who are factually guilty. Thus, the same fundamental trade-off between type 1 and type 2 errors exists under plea bargaining as at trial.

Cost shifting
In the context of civil litigation, it is often argued that a shift to the British (or loser pays) rule for legal costs might promote settlement, thereby saving judicial resources (Cooter and Rubinfeld, 1989). However, no such proposal has apparently been made for criminal litigation. This is probably true for several reasons. First, theory predicts an ambiguous effect of such a change on the settlement rate: on one hand, a loser-pays rule will tend to make trials more likely by increasing the stakes of the case, but on the other, it will tend to promote plea bargains by making trials riskier affairs (assuming litigants are risk averse). Thus, there is no necessary saving in litigation costs. Second, as noted, promoting plea bargains (as opposed to civil settlements) is not necessarily socially desirable given the possibility that some innocent defendants might be induced to plead guilty. Finally, many (perhaps most) criminal defendants cannot afford to pay their own legal costs, let alone the possibility of having to pay the state’s costs in the event of a conviction.

Empirical evidence on plea bargaining
Empirical evidence on plea bargaining confirms many of the predictions emerging from the above theory. For example, Elder (1989) found that plea bargains are more likely when the outcome of a trial is more certain, reflecting the above prediction that when the prosecutor and defendant assess the same probability of conviction, a bargain is always possible. He also found that cases involving higher stakes (more severe punishments) are more likely to go to trial, a prediction of both the mutual optimism and asymmetric information models. Finally, Elder found that plea bargains lead to shorter sentences compared to convictions at trial. Although this was not predicted unambiguously by the model (see equation (5.2)), it is consistent with the conventional notion of sentence discounts.

A comparative perspective
In contrast to its prevalence in the United States, plea bargaining is (in principle) prohibited by the inquisitorial (civil law) systems of continental

---

6 A related issue is the effect public financing of the legal costs of indigent defendants has on plea bargaining. See Garoupa and Stephen (2007).
7 Also see the studies by Landes (1971), Rhodes (1976), and Forst and Brosi (1977).
In particular, prosecutors are theoretically required to try all cases on the maximum possible charge under the principle of compulsory prosecution. Since trials cannot be avoided by guilty pleas, the problem of allocating scarce resources across a large caseload has been addressed by these systems in other ways. One is by simplifying trials and weakening some procedural protections of defendants’ rights compared to the US system, thereby lowering trial costs (Langbein, 1979). Another is by allowing prosecutors to dispose of less serious cases by cursory means that amount to little more than thinly veiled plea bargains.8

From a normative perspective, the question is whether plea bargaining or compulsory prosecution better promotes the goals of criminal procedure. In this vein, Adelstein and Miceli (2001) have argued that the two systems can be understood as efficient responses to a different set of values. Consider first the aversion to type 1 errors (punishment of the innocent). Given the reality that some wrongful convictions are unavoidable, plea bargaining actually lowers the cost of type 1 errors by allowing innocent defendants, against whom there is strong evidence, to plead guilty to a lesser charge. Easterbrook (1983, p. 320) defends this practice, arguing that there is ‘no reason to prevent [an innocent] defendant from striking a deal that seems advantageous. If there is injustice here, the source is not the plea bargain. It is, rather, that innocent people may be found guilty at trial.’

Conversely, if society places a high value on convicting the guilty (avoiding type 2 errors), then we would expect it to favor compulsory prosecution precisely because it does not allow the guilty to seek to reduce their punishment by pleading guilty to a lesser charge. Consistent with this view is that continental systems typically do not give defendants the right against self-incrimination, a right which, as we will argue below, actually helps make innocent defendants’ denials of guilt more believable. Taken together, these conclusions imply ‘that plea bargaining is more likely to evolve in systems that emphasize the protection of innocent defendants,

---

8 Garoupa and Stephen (2007) note that some civil-law countries (for example, France, Italy, Poland, and Argentina) have in fact explicitly introduced plea bargaining, despite severe criticism by traditionalists. An empirical analysis of this experiment in Italy by Boari and Fiorentini (2001), however, shows that plea bargaining has failed to produced the promised cost savings. Indeed, after five years, no more than 8 percent of criminal dispositions were achieved by plea bargains (compared to about 90 percent in the United States). The authors attribute this to institutional aspects of the inquisitorial system, and the reluctance of Italian prosecutors to adopt the new procedures.
and systems that stress punishing the guilty are more likely to be able to sustain a regime of compulsory prosecution’ (Adelstein and Miceli, 2001, p. 60).

The bail system
A fundamental question facing the procedural system is what to do with criminal defendants between their arrest and trial (or guilty plea). The trade-off is between a defendant’s right to freedom (based on the presumption of innocence), plus the cost of detent on one hand, and the cost to society should the defendant flee and possibly commit further crimes on the other. The solution is the bail system, which requires defendants to post a monetary bond as a condition for remaining free. An economic analysis of this system views the amount of bail as being determined by the above trade-off (Landes, 1973).

One criticism of the bail system is that it favors wealthier defendants who can more easily raise the necessary funds. Although bail bondsmen help to mitigate this problem, one suspects that the resulting agency problem increases the flight risk of defendants using bondsmen. Another criticism is the claim that courts set higher bail rates for minority compared to white defendants, holding the flight risk constant. Ayres and Waldfogel (1994) provide some evidence that this type of discrimination occurs.

Finally, the Bail Reform Act of 1984 made it easier to detain defendants judged to be especially dangerous. Passage of this act during a period of high crime rates in the United States is consistent with the above theory to the extent that Congress judged the increased cost of detention as being outweighed by the social interest in avoiding further crimes. The result, however, is an increase in the expected punishment for both guilty and innocent defendants (Posner, 2003, p. 579), which we will see below has an ambiguous effect on deterrence.

The standard of proof at trial
The standard of proof for conviction at trial represents a threshold for the probability of guilt beyond which the court convicts the defendant. Essentially, it represents a decision rule for transforming a probabilistic assessment of the defendant’s guilt into a dichotomous verdict. In this section, we examine the choice of the optimal standard of proof, $\theta$, in light of the goal of minimizing the cost of type 1 and type 2 errors. We also

---

examine how this approach sheds light on the differing standards of proof in criminal and civil trials.

As noted above, in the context of the model depicted in Figure 5.1, the probability of a type 1 error (false conviction) is \(1 - F_I(\hat{\theta})\), while the probability of a type 2 error (false acquittal) is \(F_G(\hat{\theta})\). Suppose society attaches a cost of \(c_1(s)\) to the former and \(c_2(h)\) to the latter, where \(s\) is the sanction imposed on conviction, \(h\) is the social harm caused by the crime in question, and \(c_1\) and \(c_2\) are increasing functions. Thus, the cost of false convictions is presumed to be increasing in the magnitude of the wrongfully imposed punishment, and the cost of false acquittals is presumed to be increasing in the severity of the crime that goes unpunished. (Again, this could reflect either a deterrent or a retributive motive for punishing the guilty.) Taking \(s\) and \(h\) as given, the problem is to choose \(\hat{\theta}\) to minimize the expected cost of the two types of errors:

\[
[1 - F_I(\hat{\theta})]c_1(s) + F_G(\hat{\theta})c_2(h).
\] (5.5)

The resulting first-order condition is

\[- F_I'(\hat{\theta}) c_1(s) + F_G'(\hat{\theta})c_2(h) = 0,
\]
or

\[
\frac{F_I'(\hat{\theta})}{F_G'(\hat{\theta})} = \frac{c_2(s)}{c_1(h)}.
\] (5.6)

The interpretation of this condition is that the marginal rate of transformation between the two types of errors should be set equal to the ratio of the social cost of those errors (Miceli, 1990; Yilankaya, 2002; Lando, 2006).

Conducting comparative statics on this condition yields\(^{10}\)

\[
\frac{\partial \hat{\theta}}{\partial s} > 0, \text{ and } \frac{\partial \hat{\theta}}{\partial h} < 0.
\] (5.7)

That is, the standard of proof should increase with the severity of the sanction imposed on conviction, and decrease with the severity of the harm caused by the offense in question. These results make sense since a given change in the standard of proof has opposing effects on the number of type 1 and type 2 errors. Thus, as the cost of type 1 errors increases,

\(^{10}\) This assumes that the second-order condition for a minimum is satisfied. Specifically, \(-F_I''c_1+F_G''c_2>0\).
the standard of proof should be raised in order to reduce the frequency of these errors. Conversely, as the cost of type 2 errors increases, the standard should be relaxed in order to reduce the frequency of these errors.

Although the actual standard of proof in criminal cases – beyond a reasonable doubt – does not vary based on these aspects of the case, the way it is applied in practice might. For example, it would be interesting to ascertain whether judges and juries define ‘reasonable doubt’ differently depending on what the punishment on conviction would be, or how severe the crime in question was.\(^\text{11}\) Dawson (1969, p. 201) provides some anecdotal evidence on the former point by noting that ‘there is judicial resistance to imposition of mandatory maximum sentences that seem unduly long in relation to the circumstances of the case’. Shepherd (2007, p. 558) likewise observes that ‘[n]ot wanting to impose a long sentence on a seemingly low-risk defendant, a prosecutor may decide not to prosecute, or a judge or jury may dismiss charges or acquit’.

One issue on which the model does shed light is the question of why there is a different standard of proof in civil versus criminal cases (Posner, 2003, pp. 618–20). The criminal standard of ‘beyond a reasonable doubt’ suggests a very high value for \(q\) (near certainty), which implies a strong bias against type 1 errors. This makes sense in the context of criminal cases because conviction often involves imprisonment, and, even in the case of probation or a fine, carries with it the stigma associated with a criminal record (Rasmusen, 1996). This is consistent with a high value of \(s\) in the above model.

In contrast, the relative costs of the two types of errors in civil cases are roughly equal, given that most disputes involve monetary damages, and there is no reason to believe that plaintiffs or defendants systematically place a higher value on a given dollar loss. This is consistent with the use of a ‘preponderance of the evidence’ standard in most civil cases, which sets the threshold for conviction at just over 0.50.

Rules of evidence
Rules of evidence dictate what evidence is permissible at trial and how it may be gathered. We consider two important constitutional rules aimed at protecting the rights of criminal defendants in the United States: the

\(^{11}\) Posner (2003, p. 620) notes that, when asked to quantify proof beyond a reasonable doubt, judges and juries generally choose numbers between 0.75 and 0.95. However, I know of no empirical study of the factors that determine how the particular threshold is chosen.
exclusionary rule and the right against self-incrimination. We also consider the defense of entrapment.

**Search and seizure**

The Fourth Amendment protects people against unreasonable searches and seizures by the police in an effort to acquire evidence in connection with a criminal investigation. The prohibition, however, is not complete; the use of a reasonableness standard suggests a balancing of costs and benefits in order to determine which searches are allowed and which are not (Posner, 2003, pp. 711–14). To illustrate, suppose that a person is suspected of having committed a crime. The expected benefit of a search is the product of the probability of uncovering evidence that will lead to a conviction, $P$, times the social gain from a conviction, $B$. The cost of a search, $C$, is the impaired privacy of the suspect, including any loss in the suspect’s reputation from the cloud of suspicion that inevitably falls on anyone who is the target of a criminal investigation. Based on these factors, the search is reasonable if and only if

$$C < PB. \quad (5.8)$$

Thus, searches that involve a substantial invasion of privacy (high $C$) are not allowed, whereas those that would lead to conviction of someone who has committed a particularly heinous crime (high $B$) are allowed.12

The law enforces this standard in two ways. The first is the requirement of a search warrant, which gives an impartial judge the opportunity to evaluate the inequality in (5.8) before authorizing a search. The second is the ‘exclusionary rule’, which prohibits the use of evidence at trial that was obtained unlawfully. There is much dissatisfaction with this rule, however, because many believe that it unduly benefits guilty defendants. An often-proposed alternative is to impose tort liability on the police for illegally seized evidence. The economic argument for this remedy is that it would allow a court to set the amount of liability in proportion to the cost of the invasion of privacy, thus allowing the police in effect to ‘buy’ the efficient amount of evidence from victims of the search. A counter-argument is that juries may be unwilling to award damages after the fact to defendants who were convicted based on evidence that was seized illegally.

---

12 See Garoupa (2007) for a more formal analysis of the trade-off between protecting the privacy rights of criminal suspects and crime prevention. He develops this analysis in the context of the European Convention on Human Rights, which strengthens procedural protections of criminal defendants, reflecting an increase in $C$ in condition (5.8).
Criminal law and economics

The rule against self-incrimination

The Fifth Amendment gives criminal defendants the right to refuse to incriminate themselves. What this means is that a defendant’s refusal to testify cannot be held against him in court. Again, some have criticized this right as primarily protecting the guilty, but an economic analysis shows that, under some circumstances, the right to remain silent can actually benefit innocent defendants by making their denials of guilt more believable (Seidmann and Stein, 2000; Mialon, 2005). This point can be demonstrated as follows.

Suppose that the right of silence were not available. Then, juries would be entitled to draw an inference of guilt from a defendant’s refusal to testify since truly innocent defendants would presumably always be willing to assert their innocence. Knowing this, however, guilty defendants would have a strong incentive to deny their guilt as well, despite the risk of being discovered in a lie. The result is a ‘pooling’ equilibrium in which innocent and guilty defendants adopt the same strategy, thereby making denial of guilt at trial an uninformative signal to juries. In contrast, the right to silence allows guilty defendants to refuse to testify without creating a presumption of guilt. In doing so, they separate themselves from innocent defendants, thus making the latter’s denials of guilt more believable.

A simple model helps to make this point clear. Let \( l \) be the probability that a given defendant is innocent – that is, \( l \) is the fraction of innocent defendants in the population of all defendants, and \( 1 - l \) is the fraction of guilty defendants.\(^\text{13}\) Also, assume that all innocent defendants testify (since, as noted, they have nothing to lose from doing so), but guilty defendants may or may not testify, depending on how the court is allowed to interpret their silence.

Consider first the case where there is no right to silence. Since the court would then be allowed to infer guilt from a defendant’s silence (given the presumed behavior of innocent defendants), all guilty defendants presumably would testify as well, even at the risk of being caught in a lie. As a result, however, the court would be able to infer nothing from the defendant’s mere act of testifying – that is, the probability of a defendant’s innocence would simply be equal to the prior probability of innocence, \( l \).

Now consider the impact of a right to silence. In this case, the court is not allowed to draw any inferences about a defendant’s guilt or innocence by his refusal to testify. Thus, while all innocent defendants will continue to testify, some guilty defendants will now choose to remain silent rather than take the stand and lie. Let \( g \) be the fraction of guilty defendants who

\(^{13}\) In this case, the probability of conviction, \( \Theta \), is assumed to be the same for all defendants.
choose to testify. As a result, the conditional probability that a defendant is innocent, given that he testifies, is

$$
\lambda' = \Pr(\text{innocent} \mid \text{testify}) = \frac{\lambda}{\lambda + (1 - \lambda)g}.
$$

(5.9)

Clearly, $\lambda' > \lambda$ as long as $g < 1$; that is, as long as some guilty defendants choose not to testify, the conditional probability that the defendant is innocent given that he testifies is larger than the prior probability of innocence.$^{14}$ In this sense, the right to remain silent, by allowing guilty defendants to remain silent without penalty, actually helps innocent defendants by making their denials of guilt more believable. As an example, if $\lambda = 0.10$ and $g = 0.15$, then $\lambda' = 0.43$. Thus, the right to remain silent results in a fourfold increase in the probability that a defendant’s testimony denying his guilt is believable.

**Entrapment**

Entrapment is a defense that a criminal defendant can raise based on the claim that he would not have committed the crime but for the solicitation of the police.$^{15}$ Thus, although the defendant is not technically innocent of the crime, he should nevertheless be acquitted. This argument makes sense both in terms of fairness (the defendant is not really a criminal), and also deterrence (since there is no behavior to deter). Thus, efforts to entrap otherwise law-abiding people are socially wasteful.

This argument does not imply, however, that solicitation of crimes is never a useful law enforcement tool. Indeed, if solicitation focuses on apprehending individuals who otherwise would have committed the crime but in circumstances where it would have been much harder to detect them, then the strategy substantially lowers the cost of law enforcement. The problem, of course, is proving in court that the individual who was nabbed by a police sting was pre-disposed to commit the crime anyway. This raises the same issues of proof discussed above regarding the determination of guilt.

**Appeals**

We have argued that the primary function of criminal trials is to minimize the cost of legal errors. The fact that such errors are unavoidable, however,

---

$^{14}$ Note that criminal penalties for perjury enhance this conclusion by making it costly to lie under oath.

provides an important economic rationale for the appeals process (Shavell, 1995; Posner, 2003, pp. 600–602). The structure of appeals in the Anglo-American legal system involves losing defendants having the right to seek redress from a higher court. One might ask whether this is preferred to a system in which a higher-level court chooses which cases to review. The advantage of litigant-initiated appeal is based on the notion that litigants (and their counsel) are presumably in a better position to observe when the trial court’s decision was erroneous (based on their private information), resulting in a filtering of cases at the appeals court level. Specifically, litigants whose cases were decided in error are more likely to be willing to incur the cost of appeal because they have a greater chance of winning. This sorting is socially desirable because it limits the work of appeals courts to cases that were more likely to have been incorrectly decided. In contrast, if appeals courts selected the cases that they would hear, this sorting would be less effective because the higher court lacks the case-specific information available to litigants.

The preceding suggests that appeals courts can be an effective means of correcting legal errors, but this gain must be weighed against the cost of adding an extra layer (or layers) to the legal process. Appeals are therefore only socially desirable if the expected gain from error correction (which in the criminal context generally involves reversal of type 1 errors) exceeds the additional administrative cost. To illustrate, suppose that legal error at the trial level, if uncorrected, imposes a cost of $500,000 on society. Also, let the cost of appeal be $150,000, and let the probability that the appeals court corrects the error be 60 percent. In this example, appeal is socially desirable because the expected gain of $300,000 = (0.60)($500,000) exceeds the cost of $150,000.

Implications for deterrence
To this point, the discussion has focused on minimizing the cost of legal error plus the cost of administration as the primary function of the procedural system, but the system also affects deterrence through a variety of avenues. First and most obviously, the procedural system determines the punishment of those defendants found guilty, whether at trial or by plea. Thus, the particular goal of criminal punishment embodied in the system will affect the behavior of potential offenders according to Becker’s (1968) rational offender model.

The appeals process also serves a lawmaking function, as common-law rules are promulgated primarily through appellate decisions. See Cooter and Rubinfeld (1989, pp. 1091–94).
Miceli (1991) explicitly compares the goals of deterrence and fairness in the Becker model with legal error.\footnote{17 Also see Harris (1970) and Polinsky and Shavell (2000b).} His results show that these goals are complementary in some respects, and contradictory in others. Specifically, more severe punishments deter crime and promote retribution (appropriate punishment of the guilty) up to a point, but they also potentially reduce fairness when punishments are perceived to be excessive,\footnote{18 The problem of excessive punishment becomes especially problematic given the need in the Becker model to scale up sanctions to reflect the uncertainty of enforcement. As a result, those offenders who are caught are in effect punished for the harm that they caused as well as for the harms caused by others who escaped detection. This prescription, which is a key element of the model of deterrence, clearly violates proportionality and also potentially runs afoul of the Eighth Amendment prohibition of cruel and unusual punishments (Adelstein, 1981).} and when the risk that truly innocent defendants are being punished becomes high. Indeed, as we noted above, there is anecdotal evidence that judges and juries are reluctant to convict defendants in the face of mandatory sentences that seem unduly harsh in relation to the circumstances of the case.

The role of the prosecutor in determining criminal punishment is also substantial since, as we have seen, the vast majority of criminal convictions are the result of plea bargains, and it is the prosecutor who ultimately decides whom to charge, and what charge to bring. Prosecutors use this discretion to conserve on limited judicial resources, but also as a check on legislatures. For example, a prosecutor who believes that the legislatively determined sanction for a particular crime is overly harsh may offer the defendant a lesser penalty in return for pleading guilty to a reduced charge (Reinganum, 1988; Miceli, 1996; Shepherd, 2007). Bjerk (2005) provides empirical evidence for this type of behavior in response to three-strikes laws for repeat offenders. Specifically, he found that prosecutors systematically react to such laws by charging offenders covered by the laws with lesser, uncovered offenses.

Apart from determining the actual punishment imposed on convicted defendants, the procedural system affects deterrence by reducing legal error. Generally, error in the determination of guilt mitigates the deterrent effect of criminal punishment by reducing the expected cost of committing a crime, or, equivalently, by reducing the benefit of being innocent (Posner, 2003, pp. 618–19). To see how, note that the expected sanction from committing a crime is \((1 - l)qGs\), where, recall, \(1 - l\) is the probability that the true offender is apprehended, \(qG\) is the probability of convicting a guilty defendant, and \(s\) is the sanction on conviction. The fact that there is legal error, however, means that there is also an expected cost of
not committing a crime. This is given by $\gamma \theta_s$, where $\gamma$ is the probability that a given individual is wrongly apprehended for a crime, and $\theta_f$ is the probability that an innocent person is convicted. The net expected cost of committing a crime is therefore $(1 – \lambda) \theta_s – \gamma \theta_s s = [(1 – \lambda) \theta_s – \gamma] s$. The deterrent effect of criminal punishment will therefore be increasing in both the effectiveness of the police in apprehending the true offender, $(1 – \lambda) – \gamma$, and in the effectiveness of the procedural system in sorting the innocent from the guilty, $\theta_s \theta_f$, holding the actual sanction, $s$, fixed.

Conclusion
This article has provided an economic perspective on criminal procedure, defined to be the rules governing the treatment of criminal defendants from the time of arrest up to the rendering of a final verdict. The primary function of this procedure is to arrive at the most accurate verdict at the least possible cost. Because of the uncertainty about a defendant’s guilt, however, there is an unavoidable conflict between the goals of punishing the guilty (whether for purposes of deterrence or retribution) and exonerating the innocent. Although popular perception of the criminal process often views the numerous safeguards of defendants as protecting the guilty, they in fact reflect a strong aversion to wrongfully convicting the innocent (type 1 error). The price for this protection is that some guilty defendants will escape punishment (type 2 errors). Ongoing policy debates regarding issues like judicial discretion in sentencing and evidence admissibility reflect this fundamental tension.

References

Note that $\gamma$ generally does not equal $\lambda$. Whereas $\gamma$ is the probability that a particular innocent person is apprehended, $\lambda$ is the probability that someone who is innocent is apprehended. Thus, we would expect $\gamma$ to be considerably smaller than $\lambda$. See Wittman (1974).


144 Criminal law and economics