1981

The Plea Bargain in England and America: A Comparative Institutional Approach

Richard Adelstein, *Wesleyan University*

CHAPTER TEN

The plea bargain in England and America: a comparative institutional view[1]

Richard P. Adelstein

10.1 Criminal price exaction and the development of procedure

From its earliest beginnings the common law has treated criminal punishment as a kind of restitution, a repayment by the offender of the damages caused by his unlawful behaviour[2]. Ours has long been a legal order of retributive punishment tempered by the norm of proportionality, one that seeks to adjust the punitive sanction to the 'gravity' of the offence and thus to exact an eye, but only that, for an eye. This element of proportionality is of central importance, for it suggests that the unconditional deterrence of all illegal activity is not in fact the fundamental organizing principle of Anglo-American criminal justice. Were the purpose of penal sanctions simply to deter all criminal behaviour, regardless of the harm it might do or the circumstances that might surround it, we would expect even trivial offences to be punished as severely as possible. Such a penalty structure would certainly discourage all but the most determined of wrongdoers, even where the social costs associated with their acts were very small, for cases in which the rewards or satisfactions derived by the offender from the act would exceed the great personal hardship imposed by punishments of this kind would be rare indeed.

The institution of proportional punishment, however, appears to contemplate a very different end. Where the personal cost the offender is forced to bear in the form of punishment is meant to reflect only the actual extent of harm his act has imposed upon the society at large, the law can be seen as tolerating (or encouraging) 'efficient offences' — those in which the material and psychic satisfactions of the act are reckoned by the offender to be greater than its full social costs as perceived by those who must bear them. To illustrate, imagine Jefferson sorely irritated, deliberating whether to punch his boss Hamilton in the nose. To simplify the exposition, let us assume that the total cost inflicted by such a blow upon Hamilton and those in society who would feel physically threatened or morally offended by it is £500, and that this sum would, with perfect certainty, be imposed upon Jefferson in the form of a fine should he commit the assault. The revenue from the fine would be transferred in full to Hamilton and the other, indirect victims of the crime. Jefferson is thus faced with the choice of committing the

crime and bearing its £500 cost, or obeying the law and forgoing the satisfaction that the blow might bring. If Jefferson decides that the intangible rewards of breaking the law would be greater than the cost of the fine, he would 'trade' with his victims by punching Hamilton and paying the attendant price.

Moreover, one could assert that, when Jefferson's welfare is included in the analysis along with that of his victims, society as a whole is made better off by his decision to break the law. If Jefferson could have been prevented from perpetrating the assault, social welfare would have been increased by the equivalent of £500 (because the social costs generated by the punch would not have been incurred) but it would, at the same time, have been decreased by the equivalent of more than £500 (since, by hypothesis, the satisfaction Jefferson forgoes by obeying the law exceeds the disutility imposed by the fine). But when the 'trade' is completed, Hamilton and the indirect victims are fully compensated for the harm done to them, and are thus equally well off, and that part of society represented by Jefferson is better off than it would have been without the assault. Where punishments are determined in this way, the aim of the criminal process appears not to be unconditional deterrence, but rather (much like the law of tort and contract) the 'internalization' of the costs of criminal activity, forcing offenders to pay those costs in full rather than allowing their imposition upon others without recompense. The incidence of crime is thus controlled, but not eliminated; inefficient offences, like goods for which no buyer is willing to pay the full cost of production, are effectively deterred, but efficient offences continue to be committed and, in the sense outlined here, each such offence represents a net increment in social welfare.

In a series of earlier essays (Adelstein, 1978b; 1979a, b; 1981), I have developed the positive and normative aspects of this 'price exaction' model of the criminal process in some detail and applied it to the evolution of procedural rules governing negotiated guilty pleas and the imposition of capital sentences under American federal law. A central postulate of this analysis is that the characteristic feature of criminal acts is the element of 'moral cost' associated with them. The widely felt sense of outrage and moral opprobrium created by particular kinds of behaviour and their accompanying states of mind that distinguishes them from other acts that might also entail damage to property or physical or psychic injury to the person[3]. These moral costs are conceptually distinct both from the physical and financial sufferings endured by the direct victims of criminal acts and from the economic dislocations that result from the efforts of other individuals to protect themselves against similar victimization. They are primarily borne by individuals with no direct relationship to the offence itself or the parties involved in it, and result from personal feelings that the set of shared values from which the social fabric is woven and which preserve the peace have been unacceptably violated. A consequence of each individual's

sense of right and wrong, they reflect the indignation created by acts that breach established and accepted moral codes of behaviour. Moreover, it is the offender's act itself and the *mens rea* that motivates it that impose these costs, and not necessarily the severity of the result or the actual harm done in economic and physical terms. Incompleted attempts and inchoate offences are thus treated as criminal acts and punished as such regardless of their ultimate outcome, and 'victimless' crimes may be understood as acts that generate *only* indirect moral cost rather than the combination of material and moral cost that results from offences in which there is an identifiable direct victim[4].

Given the existence of these costs, the institutions of criminal justice are faced with the extremely difficult tasks of assessing the full extent of social cost imposed by specific criminal acts, apprehending and convicting their perpetrators, and exacting a roughly equivalent 'punishment price' from them. But it is clear that the net social benefit that is, in principle, associated with efficient offences can be realized only if the substantial costs of actually completing the price exaction are smaller than the full social cost imposed by the original offence; nothing is gained if £100 in court costs must be expended to effect through price exaction the internalization of £50 in social damages. This economic link between the criminal law and the procedures through which it is enforced suggests that the criminal process must continually search out less costly means of exacting the punishment price if its ability to internalize the costs of criminal activity satisfactorily is to be preserved.

The initial vehicle for this adaptive response to the costliness of price exaction is the self-interested behaviour of the individuals directly involved in the compensatory transaction, the accused defendant and the prosecuting authority acting on behalf of those injured by the offence. Where material expenditure is concerned, it is these litigants who are best situated to assess the economic costs associated with the operation of the price exaction mechanism, and organizational forms within the criminal process have evolved that provide them with a continuing incentive toward procedural innovation in an effort to reduce these costs. Just as in the case of the criminal act itself, however, the process of completing the price exaction through the imposition of the punishment price generates real but nonmaterial costs, which are spread over a large class of cost bearers not immediately party to the price exaction procedure. But the organization of the prosecutorial function is such that these indirect costs cannot be fully accounted for within the structure of incentives and controls available to the litigants themselves. As a result, procedural forms developed in response to the material costs of price exaction may well impose nonmaterial social costs in excess of the economic savings involved. In both the English and American criminal processes the task of monitoring these innovations and the indirect social costs associated with them has devolved primarily upon the

appellate courts, which retain the power to encourage or foreclose various approaches on the basis of judgments made generally at the behest of defendants who claim to have been unfairly treated. A great deal can thus be learned from the careful study of such judicial outcomes within the context of the price exaction model.

Our purpose here is to add a comparative dimension to the price exaction framework by applying it to the continuing evolution of a system of induced guilty pleas in the English criminal process and relating the insights thus gained to a parallel analysis of the more explicitly developed American 'plea bargain'. Though the two criminal processes themselves are characterized by historical but significant structural differences, the adaptive organizational responses that the pressures of very similar modern environments have drawn from them bear striking resemblance to one another in terms of both form and purpose. The power of the institutional analysis lies in its ability to illuminate the nature of this evolutionary process and the implicit value judgments expressed in its outcomes. In this way, I hope, it can contribute much to the emerging English discussion of the propriety and necessity of negotiated guilty pleas.

10.2 The negotiated plea in the price exaction framework

10.2.1 Incentives toward trial avoidance: economic cost considerations

The costliness of completing the compensatory transaction itself is a central consideration in the organization of both the civil and criminal processes, for the institutional rules that apportion these costs between the interested parties largely determine which and how many cases of unlawful cost imposition will in fact be mediated by the price exaction mechanism. An example drawn from the civil side will illustrate. Suppose Blackstone negligently breaks a fine inkwell, worth £100, owned by Coke. Blackstone ungallantly refuses to compensate Coke for the loss, and Coke envisions two possible avenues for recovering the inkwell's value. He could, at nominal immediate expense but with some possibility of discovery and apprehension, break into Blackstone's home and take £100 from his cash box. Alternatively, with considerable expenditure of time and money but, let us assume, perfect certainty of success, he could bring suit against Blackstone in the local court.

Now, if Coke's behaviour is 'rational' in the economic sense, he will seek to recover his £100 in the less costly way. Imagine for the moment that the criminal law does not exist and that Coke has no compunctions about burglary. Then, even if Coke is aware that his burglary would impose economic and moral costs upon the society at large, he has a clear incentive to ignore these external effects, commit the burglary, and thus avoid the costs of bringing suit. Without the criminal law, there is no sufficiently inexpensive way for these indirect cost bearers to hold Coke accountable for *their* losses. But, of course, a principal purpose of the criminal law is to force

Coke to account for just such indirect costs in planning his activities. By imposing a penalty sufficiently severe that, even when discounted by the probability of discovery and conviction, Coke's expected costs in a recovery by theft exceeded £100, the law could effectively foreclose that option and leave him only the alternative of bringing suit. If the costs of this latter course (which in the United States but not in Britain generally include his attorney's fees in full regardless of the outcome) also exceeded £100, we could safely expect Blackstone's original tort to go unchallenged. From another perspective, if we believe *a priori* that the bearer of these recovery costs *is* 'rational', the observation that his damages remain uncompensated would create an inference that Coke has reckoned the recovery costs as greater than the value of the inkwell.

Although the organizational arrangements involved are substantially more complex, and the information necessary for such precise determinations vastly more difficult to obtain, a similar set of decisions must be made in the criminal process. The process of price exaction is itself a costly enterprise, and the magnitude of this cost in individual cases or in entire classes of offences is a principal determinant of whether or not the criminal sanction will in fact be applied. Most obvious, of course, is the economic expenditure required to apprehend and convict offenders. Where these resource costs themselves exceed the full social cost imposed by a particular criminal act, there is, just as in our tort example, a clear rationale for leaving the law unenforced in that case, a point recognized long ago by Oliver Wendell Holmes (1963, pp. 76–77). Thus, for example, we observe the highly sporadic enforcement of traffic laws and petty misdemeanours, or shifts in patterns of enforcement as once 'serious' offences become, over time, more generally tolerated forms of behaviour, as in the case of the personal use of marijuana. Where feasible, moreover, we observe as well the evolution of institutional forms, such as the withholding of income taxes from wages, where the prosecution of individual offenders would be uneconomic in this sense but where the anticipated aggregate costs of illegal behaviour are very large.

Like the criminal offence itself, however, the imposition of the punishment price is a source of moral as well as economic cost. These moral costs are generated whenever citizens not directly involved with price exaction procedures perceive them to be 'unfair' or 'improper'. Again, the term 'moral cost' is not meant to imply normative judgment; it is simply a positive measure of disutility imposed by various kinds of behaviour. A prosecutor might well save money and gain convictions were he to extract confessions by beating defendants, but it seems clear that this tactic would impose a substantial moral cost, either because it would be perceived as generating false confessions and thus punishing the innocent (cf. Posner, 1973, pp. 410–415), or for broader reasons having to do with shared values expressed in constitutional guarantees or in a general sense of fairness.

Less dramatic but of more immediate interest is the relative extent of moral cost associated with alternative procedures of price exaction, such as conviction by guilty plea rather than by trial. While decisions to litigate, settle or forgo legal action altogether in the civil process turn almost exclusively on questions of economic cost, considerations of both economic *and* moral cost must direct the corresponding decisions on the criminal side. A full adversarial trial is an extremely expensive affair, and were the economic costs of price exaction the sole concern, a decision to allow punishment only upon conviction at trial would sharply raise this cost and with it the threshold below which prosecution becomes uneconomic. But if alternatives to trial generate moral costs in excess of the economic savings they entail, the trial itself becomes the mode of conviction that lowers this threshold as far as possible and thus best responds to the *total* costs of price exaction. In this way, decisions as to whether and how to conduct the price exaction in the criminal process can be understood as closely resembling their civil analogue in form, but characterized by severe problems in gathering the information regarding both economic and widely dispersed moral costs required to define the lowest threshold and so permit the 'rational' prosecution of the greatest proportion of cases. It is the difficulty of extracting this information and bringing it to bear upon those directly responsible for the decision to prosecute that has shaped the evolution of complex and necessarily imperfect structures of incentives and authority in the English and American criminal processes.

In both Britain and the United States, the power to initiate prosecution is effectively vested in a public official who is bound by strict constraints on material expenditure but who exercises substantial discretion both in the selection of cases to be pursued and the precise charges to be brought in them[5]. In England and Wales this official is the Chief Constable of the local police, and in the Magistrates' Courts, where the vast majority of criminal cases are heard (*see*, for example, Jackson, 1972b, p. 178), it is his own police officers and assisting staff solicitors who are responsible for prosecutions in the name of the Crown. For those more serious offences that must be prosecuted in the Crown Courts, the organization of the English Bar requires that the police retain an independent barrister, briefed by the police solicitor, to represent the Crown. But here too it is the Chief Constable who determines, within broad limits, the specific charges against the defendant[6] and whose budget must bear the economic cost of prosecuting them (Jackson, 1972a, p. 164).

In American jurisdictions the police and prosecutorial functions are administered separately, although the investigatory activities of the two agencies are necessarily closely co-ordinated and prosecutorial policy strongly influences the deployment of the police. Moreover, the absence of an independent trial bar permits the public prosecutor to employ his own legal staff in the prosecution of all cases, wherever heard. But these struc-

tural differences are of far less significance than the fundamental organizational similarity between the two systems, the combination within a single authority of charging discretion and budgetary constraint in the conduct of prosecutions[7].

Prosecutorial decisions in both countries appear generally responsive to two distinct, though not irreconcilable, considerations of policy. There is, of course, the pressure of the caseload, the need to resolve as many offences as possible, preferably with a conviction of some kind, so that the various agencies of law enforcement may be seen by the community to be doing their job well[8]. At the same time, however, prosecutors are encouraged to 'individualize' the criminal sanction, tailoring dispositions to reflect the personal characteristics of each offender and the totality of circumstances surrounding each offence. This requires more than the simple maximization of prosecutorial 'output'; the prosecutor is asked instead to seek in every case the outcome he believes best serves the law enforcement interests of the whole community[9].

These concurrent objectives can be accommodated within the price exaction framework through a straightforward model of prosecutorial decision-making[10]. Let us assume the existence of a distinct prosecutorial utility function for each case in the agency's caseload. This function depicts the degree to which the prosecutor's objectives are satisfied in each case and relates the satisfaction or utility he derives from each case to the punishment price exacted from the defendant. As suggested by the individualization objective, we posit the existence of a particular sentence, say P_O, which represents the outcome that, in the prosecutor's necessarily subjective view, equates the punishment price to the social cost generated by the offence in question. The problem of caseload pressure motivates the assumption that, for punishments less than P_O, prosecutorial utility rises with increasing punishment, although at a decreasing rate. That is, each successive unit of punishment imposed upon the defendant adds a smaller positive increment to the prosecutor's utility. In this way, the relative importance of the conviction itself, of imposing *some* measure of punishment upon the offender, is emphasized[11].

Punishment prices, of course, can be exacted only upon a conviction, be it by plea of guilty or after a full trial. Where a plea agreement of some kind is involved, and where the prosecutor is confident that it will be honoured by the court at the time of sentencing, the punishments in the individual utility functions are known to him with substantial certainty before the fact. At trial, the uncertainty of conviction confronts the prosecutor with an *expected* utility instead, which depends both upon the sentence anticipated upon conviction and the probability that the defendant will be found guilty. In either case, however, the prosecutor can increase his *a priori* utility by investing some part of the material resources at his disposal in the pursuit of each case. Where trials are necessary, expenditures for case-strengthening

efforts, such as witness interviews, legal research and searches for additional evidence, can be seen as ways of purchasing positive increments in the probability of conviction and, through it, expected utility. Where guilty pleas are concerned, the effect is much the same. Given that the defendant's willingness to agree to plead guilty is highly responsive to his estimate of conviction probability, a prosecutor who increases this probability in a given case clearly strengthens his bargaining position[12].

It thus seems reasonable to suppose that the prosecutor will attempt, given the sequential nature of the cases before him and his inevitably limited ability to gather and employ the requisite information, to allocate his scarce resources to these cases so as to maximize the sum of the individual *a priori* utility functions, inducing favourable guilty pleas where he can and generating increased conviction probabilities where defendants insist upon full trials. As a result, the prosecutor will, wherever possible, select of his own volition that mode of conviction which entails the smallest *economic* cost of price exaction, just as did Coke, the private decision-maker in our tort example[13].

Suppose, for instance, that in a particular case the prosecutor, at a cost of £200 in preparation expenditures and barrister's fees, can create a 50 per cent probability of conviction at a trial in which the anticipated sentence is twenty years' imprisonment, so that the expected punishment involved is 10 years in prison. If the expenditure of £125 on preparation would persuade the defendant to plead guilty in return for a certain punishment of ten years, then even a prosecutor neutral to the risks of trial will clearly choose the cheaper mode of price exaction and save the remaining £75 for the pursuit of another case. Further, if the prosecutor is risk-averse with respect to punishment, as our behavioural assumptions imply, he will continue to choose the plea for some range of punishment *less* than the expected trial sentence of ten years rather than submit to the vagaries of the trial[14].

Although it captures the essential elements of the decision to prosecute in both England and the United States, a particularly interesting problem arises when this abstraction is considered alongside the actual organizational arrangement of English and American prosecutions. At issue is the relative diffusion of decision-making authority in the two systems. In the American criminal process, power over all the necessary legal determinations encompassed by the model is effectively concentrated in the office of the public prosecutor. Apart from his virtually unchecked discretion with respect to charge, statutorily defined constraints on the trial court's choice of sentence for various offences and the deference traditionally shown by the trial court to his sentence recommendations[15] combine to give the prosecutor an effective power to fix, and thus to negotiate, the actual sentence to be imposed upon a plea of guilty. Where the distance between the prosecutor's wishes and his ability to see them fulfilled is this short, our simple model can be applied almost without qualification.

The Chief Constable's reach, however, is not so great. While he does control the conduct of cases directly in the Magistrates' Court, he must rely upon the services of an independent and potentially 'uncooperative' barrister in the Crown Court. Moreover, although jurisdictional limitations empower magistrates to impose sentences of no greater than one year's imprisonment[16], Crown Court judges sentence quite freely at common law and under statute and are unencumbered by prosecutorial recommendations. For serious offences, then, the powers required to implement the preferences of our abstract prosecutor are in practice vested in three individuals, the Chief Constable himself, the prosecuting barrister and the sentencing judge. For our model to be of positive value in such an environment, the degree of implicit co-ordination between these agents, or the confluence of personal interests amongst them, must be very great indeed.

This is in fact precisely what we observe. The descriptive literature on induced guilty pleas in England repeatedly cites the close, 'fraternal' working relationship that exists between the Bar, from which counsel for both sides are drawn, and the judiciary[17], and the common training and outlook they share (Thomas, 1978, pp. 175–176). The 'Liberal Bureaucratic' values that pervade the entire English criminal process[18], and effectively combine an apparently genuine concern for individualized justice with a conscious sensitivity to matters of economic cost and administration, are very much like those held by American prosecutors (Miller, 1969, pp. 154–172) and by the omnipotent prosecutor in our own behavioural model. They permit the smooth operation of a system of induced pleas without centralized authority despite the ostensibly adversarial nature of the criminal process and the independence of the judge.

It is the pervasiveness of these values that makes possible the large scale inducement of guilty pleas upon which both the English and American criminal processes have come to depend. Certainly, the structural differences between the two systems have required that superficially different means to secure these pleas be developed. In both cases, the trial judge is forbidden to participate actively or directly in negotiations regarding the plea[19] and retains the formal power over sentence. But the American prosecutor's ability to fix the punishments of those who plead guilty enables him to bargain explicitly and concretely with defendants over the actual severity of their sentences, rendering the court's sentencing authority largely nugatory. The organization of English prosecutions makes overt bargaining of this kind impossible, for the trial court's power to sentence is real rather than nominal. Nevertheless, pleas *are*, within the bounds established by the Court of Appeal, systematically induced in two ways[20]. The prosecutor's charging discretion permits explicit negotiation with the defendant over the specific charges to which he will plead guilty; the defendant is technically acquitted of whatever charges he has not so admitted[21], and if the court

accepts a plea to a lesser offence in such cases, its sentence must be appropriate to this offence rather than a greater one that might originally have been charged[22]. Alternatively, the court may reduce the sentence of those who plead guilty; where such 'discounts' are routine, predictable and made known to defendants prior to plea, they amount to implicit but real plea agreements[23]. With respect to the inducement of pleas, then, structural considerations are matters of form rather than substance. The essential element of the induced plea, the effective grant of a more lenient sentence to defendants who agree to plead guilty, is as well woven into the English criminal process as it is into the American.

10.2.2 Burdening the right to trial: the defendant's choice and the problem of risk preference

We turn now to the problem posed to the defendant by this organizational arrangement. For simplicity of exposition, let us assume that the cost of legal representation and case preparation actually borne by the defendant can be neglected[24]. Then the defendant will agree to a guilty plea if he perceives the cost to him of the sentence received upon the plea as less than the expected disutility of the trial prospect and its associated sentence. That is, he will plead if

$$L(S_p) < qL(S_v)$$

where L is a function that measures the defendant's disutility for any given punishment; S_p is the sentence associated with a guilty plea; S_v is the sentence expected to result from a guilty verdict at trial; and q is the defendant's *a priori* estimate of the probability of conviction at trial.

A number of English writers (*see*, for example, Baldwin and McConville, 1977, pp. 108–109; Bottoms and McClean, 1976, pp. 234–235) have argued that the imposition of a harsher sentence upon a defendant found guilty at trial than upon one who pleads guilty amounts, *per se*, to penalizing the exercise of the right to trial. But this strong position rests upon an inapposite comparison of sentencing outcomes in two cases observed at the conclusion of all proceedings in each, after the defendants involved have made their choices with respect to plea and the consequences of those choices have been revealed. The decision to exercise any right cannot be burdened *ex post*; as our own formulation suggests, the relevant comparison is between two defendants *before* the decision by either as to plea, while the choice is still open to them. For this perspective, particular defendants may indeed find that a true 'price' has been placed on the right to trial, but the questions of when this is the case, and just how great the burden has been when it is, are rather more difficult.

Consider first a risk-neutral defendant, one to whom the disutility of a particular sentence rises in direct proportion to its length, so that each successive unit of punishment adds an equal increment of personal cost. Suppose he knows that his chances of being convicted at trial are 50 per cent and that the sentence he will receive if convicted is 10 years, so that viewed *ex ante*, the expected sentence at trial is five years' imprisonment. If the prosecutor offers this defendant a certain five year sentence in exchange for a guilty plea, the choice between bargain and trial is a matter of indifference to him because the *a priori* costs associated with each are the same. Even if the defendant goes to trial, and loses, and the judge, pronouncing the 10 year sentence, remarks, 'If you had pleaded guilty, I would have given you five years', there should be no question of impropriety[25]. The defendant has not 'paid a price' for the exercise of his right to trial; he has simply chosen to participate in a 'lottery' with two possible outcomes rather than accept a certain outcome of equal *a priori* value. Indeed, were the prosecutor to offer seven years rather than five, the trial prospect would be decidedly more attractive, much like paying £1 to buy a 50 per cent chance of winning £3 and an equal chance of winning nothing. Thus, no burden is placed on the defendant's choice whenever $S_p \geqslant qS_v$, even where S_v exceeds S_p, for the consideration offered in exchange for a guilty plea then merely reflects the uncertainties of litigation. The trial prospect represents at worst an actuarially fair gamble for the defendant, and the imposition of the longer sentence S_v upon conviction is simply the result of a losing play in a fair, or more than fair, game.

The problem of preserving the defendant's free choice arises when, given a particular probability of conviction, the sentence offered by the prosecutor falls below the expected value of the trial prospect. As we have seen, this might in some cases simply result from risk-aversion on the part of the prosecutor[26]. But it may instead represent an attempt by the prosecutor to shift, in effect, the economic costs of the price exaction to the defendant and so to deter him from insisting upon a trial. This latter interpretation is especially difficult to escape where the prosecutor seeks to influence the defendant's choice not by lowering S_p, but rather by 'overcharging', increasing S_v well beyond the actual social costs involved in the alleged criminal activity[27]. In either case, as the plea offered diverges increasingly from the expected sentence at trial, the defendant's attitude toward risk becomes a correspondingly more important factor in his decision than the 'objective' probability of conviction, raising the spectre of a significant incidence of guilty pleas by factually innocent but risk-averse defendants.

To illustrate, consider two defendants, Madison, a gambler, and Adams, who is more prudent, each facing a 50 per cent chance of conviction and a 10 year sentence if found guilty at trial. Madison's preference for risk implies that each successive year of imprisonment represents a positive but decreasing increment of disutility; for him, the 10 year term imposes less than twice

the disutility of a five year term. In order to seek an acquittal, Madison would reject an offer of five years' imprisonment (the expected trial sentence), and would continue to reject offers of *less* than five years until the prosecutor proposed the prison term that was, in Madison's eyes, half as painful as the 10 year sentence. He would, that is, prefer the trial lottery unless offered a bargain considerably better than the actuarial value of the trial prospect. Adams, on the other hand, regards the 10 year term as more than twice as costly as the five year sentence, for the pain of confinement rises at an increasing rate for each additional year. His risk-aversion would lead him to accept even sentences greater than the expected trial sentence of five years, perhaps seven or eight years, to avoid the possibility of the ten year term.

Now, while Madison would reject a prosecutorial offer of five years which Adams would accept without hesitation, the sentencing discount associated with guilty plea might be so great, say a two year term, that even Madison would be persuaded to relinquish his right to trial and agree to plead guilty. If Adams were so risk-averse that he would have accepted a seven year term rather than face the risk of 10, would an offer of two years be so powerful an inducement as to overbear his will, entirely apart from his factual guilt or innocence? For an extremely risk-averse defendant, moreover, even a plea agreement representing only the expected sentence at trial poses this problem. Given the sense of voluntary behaviour underlying models of economic 'rationality', none of these agreements would be seen as coerced, for each party was at least as satisfied with the agreement as with the alternatives. But the law's view has generally been more subtle[28], and in both the English and American courts, it is on this ground of volition and coercion that the propriety of induced pleas has been argued.

10.3 External aspects of the price exaction procedure: moral cost considerations

10.3.1 The role of appellate courts

The organizational forms discussed thus far are, in their nature, able to account only partially for the costs involved in imposing the punishment price. The combination of budgetary constraints, broad charging discretion and, in the English case, the sentencing leniency routinely shown to those who plead guilty, provide the prosecution with both the incentive and the opportunity to reduce the purely *economic* costs of price exaction as much as possible. But once the prosecutor decides upon a particular course of action, the budgetary mechanism alone is insufficient to bring the *moral* costs associated with various price exaction procedures to bear upon his selection of one mode of conviction or another. Thus, even if negotiated or induced guilty pleas were to generate moral costs, specifically those created by the

risk of systematically erroneous convictions (given the relaxation of safeguards inherent in the full criminal trial), that outweighed the economic savings involved, their incidence might remain very high because nothing compels the prosecutor to account for moral costs that do not impinge upon his budget. In this sense, the prosecutor has no incentive to allocate his budget in such a way as to address all the community's interests in choosing a mode of price exaction.

It is here that the appellate courts play an important corrective role[29]. The potential existence of significant moral costs borne by members of the community at large, not party to the price exaction procedure itself, is first signalled by the claims of one or more individual defendants that their own guilty pleas were entered involuntarily as a result of the inducements or pressures confronting them. To the extent that the courts see this inherently subjective claim as general and prototypical, they attempt to 'objectify' it by testing it against their own subjective evaluation of the degree of 'unfairness' entailed in the procedures at issue, a determination based largely upon judicial perceptions of widely shared values of criminal justice and notions of 'reasonable' behaviour. Within the price exaction framework, this process can be interpreted as an effort to evaluate the uncounted moral costs of the challenged procedure and, where these are seen to exceed the savings in economic expenditure that motivated it, to foreclose this aproach and direct the criminal process to an alternative mode of conviction that more successfully addresses the full costs, both economic *and* moral, of price exaction.

It must be recognized that this process of structural evolution in the price exaction mechanism is not an 'optimizing' process. The strictly limited ability of appellate courts to make accurate assessments of the moral costs associated with various procedural alternatives precludes any suggestion of 'optimal' decisionmaking. Decisions that might be seen as optimizing under hypothetical conditions of perfect information may be rendered suboptimal or even dysfunctional given the potential for error introduced by these informational constraints. Nevertheless, the price exaction analysis does suggest that, as a *positive* matter, the judicial resolution of these questions can be taken as a rough measure of the relative magnitude of the economic and moral costs involved as perceived by those bodies charged with estimating them. They represent the implicit value judgments and assumptions of fact that underlie the common English and American preference for 'bargain justice'.

10.3.2 The negotiated guilty plea in the United States Supreme Court

Although the problem had been addressed previously and somewhat fitfully in the lower federal courts[30], the Supreme Court did not begin serious

consideration of the negotiated plea until 1968. In *United States v Jackson*[31], the Court indirectly spoke to their propriety in striking down a provision of the Federal Kidnapping Act[32] that permitted the death penalty to be imposed only after a full trial by jury, rather than upon a guilty plea. 'The inevitable effect of any such provision,' said the Court, 'is, of course, to discourage assertion of the Fifth Amendment right to demand a jury trial. If the provision had no other purpose or effect than to chill the assertion of constitutional rights by penalizing those who choose to exercise them, then it would be patently unconstitutional.'[33] The Court conceded that in this instance the government might have had another legitimate purpose in mind, mitigating the severity of the death penalty by commending the decision to the trial jury. But '[w]hatever might be said of Congress' objectives, they cannot be pursued by means that needlessly chill the exercise of basic constitutional rights. The question is not whether the chilling effect is "incidental" rather than intentional; the question is whether that effect is unnecessary and therefore excessive.'[34]

Implicit here is the suggestion that sentencing concessions that go beyond the uncertainties of litigation might be 'patently unconstitutional' as well[35], for as we have argued, the very purpose of such inducements is to discourage defendants from insisting upon costly trials. But the trial procedure is not the only source of cost that the defendant might impose upon the state. Where the defendant invokes a statutory right to appellate review of his conviction, the state is forced not only to bear the costs of contesting the appeal but, should the attack be successful, to retry or negotiate with the defendant or abandon its prosecution of the case. Just as in the case of plea negotiations, the exercise of this right can effectively be deterred by the threat of a substantially more severe sentence for those whom the state must reconvict after a successful appeal. In *North Carolina v Pearce*[36], however, the Court reinforced the *Jackson* result by holding this practice unconstitutional as well. 'It...would be a flagrant violation of [due process] for a state trial court to follow an announced practice of imposing a heavier sentence upon every reconvicted defendant for the explicit purpose of punishing the defendant for having succeeded in getting his original conviction set aside.' Once the state chooses to establish a right of appellate review, it is 'without right to...put a price on an appeal. A defendant's exercise of a right of appeal must be free and unfettered....[I]t is unfair to use the great power given to the court to determine sentence to place a defendant in the dilemma of making an unfree choice.'[38]

Yet when the Court did confront the question of negotiated pleas squarely in *Brady v United States*[39], it ignored the implications of *Jackson* and *Pearce* and upheld the general practice of plea bargaining in its most common forms against a claim that such pleas were inherently involuntary. Stressing the 'mutuality of advantage'[40] to prosecutor and defendant that underlies the negotiated plea, the Court endorsed offers of leniency 'to a defendant who

in turn extends a substantial benefit to the State.'[41] It condemned as involuntary only those pleas produced by threats of physical harm or by mental coercion 'overbearing the will of the defendant'[42]. The definition of such overbearance in terms of risk-aversion considerations is not made clear, but there is a suggestion that prosecutorial overcharging or judicial abuse of the sentencing power might render such a plea involuntary[43]. The Court made clear that it *was* sensitive to the hazards of plea negotiations. If it could have been shown that a particular bargaining tactic, or the negotiation process itself, 'substantially increased the likelihood that defendants, advised by competent counsel, would falsely condemn themselves'[44], the Court would have had serious doubts about the case.

> But our view is to the contrary and is based on our expectations that courts will satisfy themselves that pleas of guilty are voluntarily and intelligently made by competent defendants with adequate advice of counsel and that there is nothing to question the accuracy and reliability of the defendants' admissions that they committed the crimes with which they are charged.[45]

This expectation that false pleas by competently advised defendants will be rare may itself rest on a belief, shared by a former Solicitor General of the United States[46], that preliminary screening of cases by the police, the prosecution and, in some cases, the grand jury, will ensure that no defendants actually faced with the decision to accept or reject a plea agreement are, in fact, innocent of the offence in question.

While the *Brady* Court would not go so far as to acknowledge explicitly the reason for this apparent retreat from *Jackson* and *Pearce*[47], Chief Justice Burger did so almost casually the following term in *Santobello v New York*[48]:

> '[P]lea bargaining' is an essential component of the administration of justice. Properly administered, it is to be encouraged. If every criminal charge were subjected to a full-scale trial, the States and the Federal Government would need to multiply by many times the number of judges and court facilities.

As the Chief Justice suggests, moreover, the obvious reason for this reliance upon negotiated pleas is not in fact contrary to the language of *Jackson*. For the Court did not hold in that case that *all* procedures designed to deter the exercise of the right to trial were forbidden; rather, the central question is whether a particular procedure 'is *unnecessary* and therefore excessive'. The Court appears persuaded, not without reason, that where the fulfillment of the constitutional guarantee of a day in court for every defendant would dramatically increase the economic claims of criminal justice on ever scarcer

social resources, 'the promise must be tempered if society is unwilling to pay its price'[49]. Relative to the economic costs of providing a full trial for every defendant accused, of course, the actual costs to the states and the federal government of contesting appeals and reconvicting successful appellants are in practice very small. Indeed, one suspects that were the American criminal process someday to be threatened with suffocation by potential appellants as it has already been by prospective trial defendants, the procedures struck down in *Pearce* would become as 'necessary' as the very similar tactics upheld in *Brady*. But within the price exaction framework, these cases can be read as an implicit evaluation that the moral costs associated with the attempt to burden the right of appeal outweigh the economic savings it would entail, but that the vastly greater economic costs involved in the case of negotiated pleas compel the opposite result.

It would, I think, be mistaken to think that the Court has not been aware of this implicit cost accounting, or that its members experienced no uneasiness over it. In *Bordenkircher v Hayes*[50], the Court considered the propriety of a prosecutor's threat to reindict a defendant under the state's habitual offender statute, which carried a mandatory life sentence, should he refuse the offer of a plea agreement on a charge of uttering a forged cheque for $88.30. Although the prosecutor's right to invoke this statute under state law was unchallenged and there was sufficient evidence to support the indictment, the issue was posed in terms of prosecutorial 'vindictiveness', suggested by language in *Pearce*: 'Due process of law, then, requires that vindictiveness against a defendant for having successfully attacked his first conviction must play no part in the sentence he receives after a new trial.'[51]

In upholding the defendant's life sentence under the recidivist statute after his insistence upon a trial, the Court said:

> To punish a person because he has done what the law plainly allows him to do is a due process violation of the most basic sort, and for an agent of the State to pursue a course of action whose objective is to penalize a person's reliance on his legal rights is 'patently unconstitutional'. But in the 'give-and-take' of plea bargaining, there is no such element of punishment or retaliation as long as the accused is free to accept or reject the prosecutor's offer.[52]

The distinction the Court attempted to draw here is clearly a very problematic one, for it chose the wrong hypothetical defendant upon which to base its analysis. It is not the defendant that accepts the offer that is retaliated against in the plea bargain, it is those, like Hayes, that *refuse* the offer and are sentenced accordingly after trial that must pay the penalty for exercising their rights. There is no inducement to plead without the example of the Hayeses to place before recalcitrant bargainers. The Court itself seemed to

recognize the futility of distinguishing *Pearce*, for it soon conceded that 'by tolerating and encouraging the negotiation of pleas, this Court has necessarily accepted as constitutionally legitimate the simple reality that the prosecutor's interest at the bargaining table is to persuade the defendant to forgo his right to plead not guilty.'[53]

10.3.3 The induced guilty plea in the Court of Appeal

In contrast to the open, occasionally agonized, approach of the American Supreme Court, the Court of Appeal's endorsement of the systematic inducement of guilty pleas has been indirect and implicit. To be sure, the Court of Appeal has repeatedly and in emphatic terms attacked overt discussions between the trial judge and defence counsel over the precise sentencing concessions to be awarded a defendant who enters a plea of guilty. But at the same time, the Court has clearly established 'procedures which can produce results akin to those produced by direct plea bargaining' (Cross, 1975, p. 103) and taken some care to preserve those procedures during its periodic denunciations of negotiated pleas. The result, as we shall argue, has been to create a system of plea inducement in which the transfer of essential information is specifically impeded and which may, therefore, often operate so as to defeat its own purpose.

The keystone of the English structure is the well-settled principle that a defendant's plea of guilty ought in general to mitigate the sentence he receives, ostensibly in recognition of the contrition evinced by the plea (*see* Thomas, 1979, pp. 50–52). In *Harper*[54], for example, the Court reduced a sentence of five years' imprisonment after a trial in which the defendant had alleged improper behaviour by the police.

> This court feels it quite improper to use language which may convey that a man is being sentenced because he has pleaded not guilty, or because he has run his defence in a particular way. It is, however, of course proper to give a man a lesser sentence if he has shown genuine remorse, amongst other things by pleading guilty.[55]

While the wisdom of such a policy can be questioned[56], if the element of contrition is genuine and in fact the source of the defendant's plea, this 'sentencing discount' involves no logical contradiction. But this rationale was immediately weakened substantially by the Court in the companion case of *de Haan*[57]. There, the fact of the defendant's guilty plea itself, without any independent showing of remorse, was held to *require* the granting of leniency, 'for that is clearly in the public interest'[58]. In practice, moreover, both the mandate and the example of *de Haan* appear to guide the practice of sentencing judges, who routinely award the discount to those who plead guilty without even a *pro forma* attempt to establish the defendant's

motives[59]. In such circumstances, the logic of *Harper* is severely undermined, a point not lost upon defendants faced with the necessity of choosing a plea (Bottomley, 1973, p. 122).

Against this background, the Court first confronted the induced plea and its relationship to the sentencing discount in *Turner*[60]. Charged on complex issues of fact with theft, Turner at first persisted in a plea of not guilty despite the urgings of his own counsel and briefing solicitor. Just as the trial was to begin, however, he relented and changed his plea to guilty in response to his counsel's 'personal opinion'[61], formed on the basis of a private meeting with the trial judge, that such a plea was likely to result in a non-custodial sentence rather than the prison term anticipated upon conviction at trial. Turner appealed, claiming that he was thus deprived of a free choice as to plea, and the Court agreed. In remanding the case for a *venire de novo*, it offered some observations on the 'vexed question of so-called "plea bargaining"'[62].

The Court explicitly forbade judicial statements to the effect that a specific discount would be awarded on the basis of a guilty plea[63]. Yet at the same time, it reiterated the propriety of the discount itself and specifically charged defence counsel with the 'duty' to inform the accused of the fact of its existence[62]. Beyond this, 'when it is felt to be really necessary'[64], the Court appeared to encourage private discussions between both counsel and the trial judge of matters that might be 'of such a nature that counsel cannot in the interest of his client mention them in open court,'[62] including the desirability of a guilty plea to a lesser charge.

As we have seen, the overt participation of the trial judge in plea negotiations is generally proscribed in the United States as well and is, in any case, unnecessary where the sentencing discount is regularly granted and its extent can reliably be communicated to the defendant. Hence the ambiguity of the *Turner* directions, which seem simultaneously to forbid the inducement of guilty pleas and to sanctify the institutional arrangements by which it can be carried out effectively. This apparent contradiction was made explicit but left quite unresolved by the Court in *Cain*[65]. There, the trial judge sent for both counsel during the trial and made clear to them that if the appellant persisted in what he saw as a futile defence he would be sentenced very severely, but that a changed plea of guilty would be favourably received and make 'a considerable difference'[66]. Applying *Turner*, the Court found Cain's subsequent guilty plea involuntary and ordered a new trial. But the Court went on to suggest that where defence counsel was unfamiliar with the trial judge and his sentencing practice, private discussions aimed at 'obtaining guidance' as to what sentence the judge had in mind so that his client might properly be advised were indeed within the *Turner* guidelines.

> It was trite to say that a plea of guilty would generally attract a somewhat lighter sentence than a plea of not guilty after a full dress contest on the

issue. Everybody knew that it was so, and there was no doubt about it. Any accused person who did not know about it should know it. The sooner he knew the better.[67]

With *Cain*, as one commentator has aptly written,

> The law seems to have got into a very confused and puzzling state. The accused ought to know that a plea of guilty will attract a lighter sentence than he would receive after conviction on a plea of not guilty .... Yet, where precise information is available, he is to be denied it: so that, if the accused decides to plead not guilty, he will or may do so on a false premise and one which — in the circumstances of the present case — his counsel knows to be false.[68]

A terse and somewhat cryptic Practice Direction[69], issued shortly after *Cain*, did little to resolve the tensions inherent in *Turner*. Indeed, the Court has continued to disclaim the form of plea agreements while carefully preserving their substance on a series of opinions[70] that border on the disingenuous. In *Atkinson*[71], for example, the appellant was convicted at trial of handling stolen goods and sentenced to six months' imprisonment. At a hearing conducted well before the trial, the judge suggested in open court to Atkinson's counsel that 'if he decides to change his plea [to guilty] we can dispose of it all today, and he would be out in the sunshine.'[72] As the Court of Appeal itself noted, '[t]hat was a clear indication, which no doubt was faithfully conveyed to the appellant, that if he pleaded guilty there would be no question of his going to prison.'[72] But Atkinson persisted in his plea of not guilty and, as sentence was passed, the judge recalled this earlier episode:

> And I observed that I there indicated that...I had it in mind sending this man to prison. But he has not pleaded guilty, has he? And the position of a man who pleads guilty is one thing, because he can say, 'Well, I am sorry. I am showing, by pleading guilty, I am not going to put the public to further expense.' A man who is found guilty, having denied it, is in a far different position, is he not?[72]

In allowing the appeal, the Court appeared to make its position on such procedures clear:

> Plea-bargaining has no place in the English criminal law....Our law having no room for any bargain about sentence between court and defendant, if events arise which give the appearance of such a bargain, then one must be very careful to see that the appearance is corrected.... [S]uch useful devices as a pretrial review must not be used by the court to indicate to an accused man that he may be treated one way if he pleads not guilty but in another way if he pleads guilty.[73]

But remarkably, on the facts of this case, the Court also said:

> Of course the trial judge was not striking any bargain with the defence. He was indicating the difference in sentence that a man can on occasions secure in his favour by a plea of guilty.... Although the learned judge no doubt had no intention of making a bargain with the defence as to plea, it may well have appeared to the appellant that he was being offered the relief from a sentence of immediate imprisonment if he should decide to plead guilty.[74]

In this way, the Court reiterated its general approval of the sentencing principles set forth in *Harper* and *de Haan*, and went on to sanction both open and private communications between judge and counsel within 'the limits set in *R v Turner*'[73]. Where it clearly lay within the power of the Court to eliminate the sentencing discount itself and to close the channel through which, inevitably, information regarding the extent of that discount in particular cases will be sought, the Court's refusal to do either necessarily renders its pronouncements on 'plea bargaining' somewhat hollow (cf. Baldwin and McConville, 1979, pp. 202–208). However 'damaging to the face of justice'[75] the inducement of guilty pleas may be, the Court appears, like its more candid American counterpart, to have found that damage worth bearing in light of the economic savings achieved by those pleas.

Once this underlying value judgment has been made, however, the full benefit of a system of induced guilty pleas can be realized only where essential information is allowed to flow freely within it[76]. As a result, the Court's continuing refusal to make precise information about the sentencing discount available to defendants prior to their decision as to plea serves only to frustrate the purpose of the discount itself. As we have argued (*see* section 10.2.2), the defendant will relinquish his right to trial where the expected cost to him of the trial prospect exceeds that of the proposed plea agreement, that is, where

$$L(S_p) < qL(S_v)$$

In these terms, the effect of the Court's policy is to create substantial uncertainty in the defendant's mind about the value of the variable S_p. Suppose, for example, that counsel is able to tell a particular defendant that a conviction at trial would mean a sentence of 10 years, but only that 'some reduction' is likely to follow a guilty plea. When the defendant asks how great this reduction will be, counsel's reply must necessarily be seen as merely an estimate; 'four years', therefore, must be understood by the

defendant as 'four years, but possibly less'. Now, if the defendant would in fact have accepted a plea offer of six years, but gone to trial if the term offered were any greater than this, the uncertainty surrounding the decision would result in a trial (and its attendant costs) even if the judge fully intended to impose a sentence of six years following a guilty plea.

More generally, the uncertainty implied by the Court's policy requires that any estimate S_p, be it correct or not, made by a defendant prior to plea must be *increased* by some amount, say h, to reflect the possibility that the estimate is too low. If

$$L(S_p) < qL(S_v) < L(S_p + h)$$

that is, if the expected cost to the defendant of the trial prospect is greater than that of a 'certain' agreement at S_p but less than that of an 'uncertain' one, an uncertain defendant will elect a trial, even though the judge was in fact prepared to sentence him in such a way as would otherwise have induced a guilty plea. When this occurs, the potential welfare gains to both prosecution and defence that would have accompanied the guilty plea are never realized. The 'price' of an induced plea is increased, as relatively lower values of S_p must be offered to secure an agreement. The prosecution thus pays a premium because of the uncertainty forced upon defendants by the Court's policy and, if this uncertainty is sufficiently great, *no* agreements will be struck despite the desire of all sides to see them completed.

Deary[77] suggests that the 'real sense of grievance' engendered by uncertainty of this kind may require adjustment of sentence to reflect the reasonable expectations of defendants induced to plead guilty. But if, as the commentator in that case argues, '[it cannot] be treated as a decision that "plea bargains" will be enforced by the Court of Appeal,'[78] the problems created by the Court's anomalous position will persist. I have argued elsewhere that rules or policies that deny individuals necessary information in just this way can only be seen within the price exaction framework as transient (Adelstein, 1981). As a positive matter, then, I would suggest instead that *Deary* represents a first step toward the reconciliation of the Court's position regarding information transfer with its larger policy regarding induced pleas themselves.

10.4 A direction for further research

The principal purpose of this chapter has been to consider the development of the English system of induced guilty pleas and the nature of the value judgments that support it from the particular vantage point offered by the price exaction analysis. From a more general perspective, however, this comparative study can itself be seen as a kind of 'experiment', an attempt to

test the usefulness of the framework as a theoretical basis for understanding the evolution of structural and organizational forms within the criminal process (cf. Adelstein, 1981, pp. 196–198). But this first step has necessarily ben a short one, for the common sources and close historical relationship between the English and American criminal processes have generally caused their similarities to dominate their differences and thus left relatively little scope for significant comparative analysis.

The real value of the framework, I believe, lies in its potential for illuminating the parallel evolution of systems of criminal price exaction characterized by different historical experience and substantial variance in the structure of basic institutions. Consider, for example, the degree of independent authority vested in the public prosecutor in the selection and pursuit of individual cases, an element that, in principle, distinguishes the Anglo-American criminal process from the 'inquisitorial' or 'mixed' systems of Europe. Substantial and largely unchecked prosecutorial discretion in these areas is a pervasive part of the Anglo-American system, but continental law tends to favour a rule of compulsory prosecution (or 'principle of legality') in cases of serious crime, monitored by a process of judicial review of decisions not to prosecute (*see*, for example, Langbein, 1977, pp. 87–89, 100–105, 111–115). But all these systems must somehow deal with the problem of selecting which offences to prosecute, and how, when the available human and material resources are insufficient to provide every defendant with a full criminal trial. As we have seen, the Anglo-American response has been to resort to some kind of plea agreement, a course made possible only by the freedom of prosecutors and courts to adjust the charges against a particular defendant to correspond with the terms of the agreement. On the European side, the legality principle forecloses this approach, but continental systems have in fact developed organisational forms consistent with mandatory prosecution that address the problem of case selection in the face of limited resources. The German forms of *Strafbefehl* and *Opportunitätsprinzip*, for example, are qualitatively different from their Anglo-American analogues of plea agreements and substantial prosecutorial discretion (ibid., generally, pp. 87–111), and close attention must be paid to the reasons for and effects of these differences. But analysis within the price exaction framework also makes clear that they serve many of the same purposes and arise for many of the same reasons, and much can be learned from this as well.

From a still wider perspective, it is clear that, in addition to the fundamentally different approach to economic analysis it implies, comparative study of this kind must also draw heavily upon the skills and analytical approach of legal scholarship. Perhaps the most important contribution of the price exaction framework will be the essential place it creates for legal analysis and a recognition of the much larger role that must be played by legal analysis in the emerging field of economics and law.

Notes to chapter ten

1 This essay is a product of an academic year spent by the author as Visiting Fellow, Centre for Socio-Legal Studies, Wolfson College, Oxford. I am deeply grateful to the staff of the Centre for their friendship and intellectual support during this most enjoyable year, and to Wesleyan University and the British Social Science Research Council for the financial support that made it possible.

2 Pollock and Maitland (1968, p. 451) trace the English practice to the time before the line separating crime from tort had been sharply drawn:

'The deed of homicide is thus a deed that can be paid for by money. Outlawry and blood-feud alike have been retiring before a system of pecuniary compositions.... From the very beginning...some small offences could be paid for; they were "emendable". The offender could buy back the peace that he had broken. To do this, he had to settle not only with the injured person but also with the king.... A complicated tariff was elaborated. Every kind of blow or wound to every kind of person had its price, and much of the jurisprudence of the day must have consisted of a knowledge of these preappointed prices.'

3 See, generally, Adelstein (1979b). The incorporation of nonmaterial and moral effects into discussions of economic efficiency has been suggested elsewhere in the legal literature. See, for example, Michelman (1967, pp. 1173, 1214–1218); Calabresi and Melamed (1972, pp. 1111–1112); and University of Pennsylvania Law Review (1974). Compare the earlier and more traditional discussion of Cohen (1940).

4 It is most important to emphasize the positive nature of the concept of moral cost. These effects are postulated solely to capture a real social phenomenon that has been reflected in Western attitudes toward crime and punishment for centuries. But to argue that such costs exist and play a central role in the criminal process is not to imply approval or justification for a given instance of them. Thus, for example, that a statutorily illegal act by a person of one race or nationality may often generate greater moral cost than an otherwise identical act by a person of another does not provide ethical support for the intolerance that makes it so, but it may help to explain the unequal punishments observed in the two cases. More generally, the positive magnitude of these moral costs appears quite variable and highly sensitive to the specific details of each criminal act. The identities of victim and offender and the particular circumstances that surround a given offence are, in practice, principal determinants of the 'gravity' or 'seriousness' of the offence and thus of the punishment to be imposed upon the offender. It is, moreover, this individualization that eliminates the possibility of competitive forces that might otherwise facilitate the organization of explicit markets in these effects. See Adelstein (1978b, pp. 793–796).

5 Although private prosecution remains a fundamental principle of English criminal procedure, its role is largely symbolic rather than practical. In Hertfordshire, for example, during a three month period in 1969 only 81 of 9341 prosecutions (less than one per cent) were brought by private citizens. See Wilcox (1972, pp. 3–4); Jackson (1972a, pp. 155–158).

6 *Metropolitan Police Commission, ex parte Blackburn*, [1968] 2 QB 118; Williams (1956).

7 On prosecutorial discretion with respect to the charging decision, see *Powell v Katzenbach*, 359 F (2d) 234 (DC Cir. 1965), cert. denied 384 US 906 (1966); *United States v Cox*, 342 F (2d) 167 (5th Cir. 1965), cert. denied 381 US 926 (1965); Cox (1976, pp. 394–403). From the perspective of price exaction, this charging discretion is the means by which the prosecutor seeks to adjust the punishment imposed upon the offender to his perception of the largely moral costs associated with the criminal offence. Since these costs are themselves borne by a large number of persons that are not direct participants in the price exaction process, the punishment sought by the prosecutor necessarily reflects his own subjective and often imperfect estimate of these social costs. As we have suggested, the need for a 'principal-agent' relationship of this kind creates severe problems of information gathering and introduces a substantial potential for error into the prosecutorial decision-making process that is not present where costs are concentrated in specific individuals who retain the responsibility for decisions based upon them.

8 With respect to the United States, see Cox (1976, pp. 413–415). On comparable English pressures, see Jackson (1972b, pp. 80–82); Bottomley (1973, p. 108).

9 American practice is reported in Miller (1969, pp. 161–165) and Newman (1966, pp. 112–130). The English policy is considered in Jackson (1972b, pp. 88–93) and Wilcox (1972).

10 These arguments are developed more formally and in much greater detail in Adelstein (1978a).

11 For punishments exceeding P_O, we need make no specific assumptions concerning the shape of the prosecutor's utility function. Thus, the value of the function (the degree to which the prosecutor is satisfied) may fall when punishment exceeds P_O, reflecting the prosecutor's sense that P_O is in fact the most appropriate outcome. Alternatively, it may remain constant or rise slightly beyond P_O to account for the value of these more severe sentences as bargaining chips in negotiations with other defendants.

12 We consider the defendant's problem more fully in section 10.2.2. Insofar as pretrial detention may be a less costly means of 'softening' defendants and inducing guilty pleas than investigation or trial preparation, prosecutors may resort to this tactic as well. See, for example, Blumberg (1970, p. 59). 'Bail bargaining' along these lines appears to be quite common in England. See Bottoms and McClean (1976, pp. 200–204); Heberling (1978, pp. 101–103). The Court of Appeal took judicial notice of this practice in *Northam*, (1967) 52 Cr App R 97, 100.

13 Note that the remaining economic costs of price exaction, such as the cost of apprehending suspected offenders, are substantially independent of whether the conviction follows a guilty plea or a full trial.

14 This is a consequence of the assumption of diminishing marginal utility of punishment. The difference between the expected sentence at trial and the smallest acceptable certain punishment below it represents the prosecutor's 'risk premium', the value to him, in units of punishment, of removing the uncertainty inherent in the trial.

15 This deference is increasingly being institutionalized in law. See, for example, the relevant rule in the American federal courts: '(e) Plea agreement procedure... (3) *Acceptance of a Plea Agreement.* If the court accepts the plea agreement, the court shall inform the defendant that it will embody in the judgment and sentence the disposition provided for in the plea agreement.' *Federal Rules of Criminal Procedure* 11 (e) (3) (1976).

16 Magistrates' Courts Act 1952, s. 108. This sentencing constraint creates a strategic incentive for defendants to have their guilty plea entered in the Magistrates' Court wherever possible. See Heberling (1978, p. 97).

17 See, for example, Heberling (1978, p. 96); Baldwin and McConville (1977, pp. 84–85, 111). One writer has described the English system of induced pleas as a 'delicate mutual back-scratching system' Parker (1971, p. 408).

18 See Bottoms and McClean (1976, pp. 228–235): 'The Liberal Bureaucratic Model holds ... that the protection of individual liberty, and the need for justice to be done and to be seen to be done, must ultimately override the importance of the repression of criminal conduct.... [But] the liberal bureaucrat is a practical man; he realises that things have to get done, systems have to be run. It is right that the defendant shall have substantial protections; crime control is not the overriding value of the criminal justice system. But these protections must have a limit. If it were not so, then the whole system of criminal justice, with its ultimate value to the community in the form of liberal and humane crime control, would collapse. Moreover, it is right to build in sanctions to deter those who might otherwise use their "Due Process" rights frivolously, or to "try it on"; an administrative system at State expense should not exist for this kind of time-wasting.... [T]he values of the Liberal Bureaucratic Model are everywhere to be found in the actual operation, and even in some of the formal rules, of the English courts.... All these rules help to smooth administrative operation of the system, while leaving open to the defendant his formal rights — a classic statement of the Liberal

Bureaucratic position.' (pp. 229–231).

19 The American federal courts are governed by *Federal Rules of Criminal Procedure* 11 (e) (1) (1976); see also the discussions in *Scott v United States*, 419 F (2d) 264 (DC Cir. 1969) and *United States ex rel Elksnis v Gilligan*, 256 F Supp. 244 (S.D.N.Y. 1966). On the English side, see *Turner*, [1970] 2 QB 321 and *Atkinson* [1978] 2 All ER 460.

20 On induced pleas in the English criminal process generally, see Thomas (1978); Heberling (1978); Baldwin and McConville (1977); Bottoms and McClean (1976, pp. 104–134); Jackson (1972b, pp. 163–171).

21 Criminal Justice Act 1972, s.17.

22 Cf. *Kennedy*, [1968] Crim LR 566.

23 On the legitimacy of the sentencing discount, see, for example, *Harper*, [1968] 2 QB 108, a case we discuss more fully in section 10.3.3. The extent of the discount in practice is the subject of Baldwin and McConville (1978).

24 In both the United States and England, the vast majority of defendants who require representation have counsel provided for them at public expense. Compare, for example, President's Commission (1967, pp. 152–161) with Jackson (1972b, pp. 133–142). This simplifying assumption is relaxed in the model developed by Adelstein (1978a).

25 Cf. *Scott v United States*, 419 F (2d) 264, 277–278 (DC Cir. 1969).

26 Note that where this is the case, prosecutors will bargain most readily with defendants whose guilt is in substantial doubt, precisely those, it can well be argued, who most deserve the formalities of the trial procedure. Trials will thus be likely in only the most perfunctory cases, where the defendant's guilt is easily proved and the prosecutor has no reason to concede anything.

27 This practice apears to be more common in the United States than in England. Compare, for example, President's Commission (1967, pp. 9–13) with Baldwin and McConville (1977, pp. 111–112).

28 Thus, in considering a fifteen-year-old boy's station-house confession, Justice Frankfurter argued: 'It would disregard standards that we cherish ... to hold that a confession is "voluntary" simply because the confession is the product of sentient choice. "Conduct under duress involves a choice." [C]onduct devoid of physical pressure but not leaving a free exercise of choice is the product of duress as much so as choice reflecting physical constraint.' *Haley v Ohio*, 332 US 596, 606 (1948) (Frankfurter J, concurring) (Quoting *Union Pacific R Co v Public Service Commission*, 248 US 67, 70 (1918)).

29 This argument, and the empirical and epistemological issues surrounding it, are discussed in detail in Adelstein (1981, pp. 154–198).

30 See, for example, the tortuous history of *Shelton v United States*, 242 F (2d) 101 (5th Cir.), *rev'd en banc on rehearing*, 246 F (2d) 571 (1957), *rev'd per curiam on confession of error*, 356 US 26 (1958).

31 390 US 570 (1968).

32 18 USC s.1201 (1976) (amended 1972). The 1972 amendment struck out the provision at issue in *Jackson*.

33 390 US at 581.

34 Ibid. at 582 (citations omitted).

35 Indeed, there were suggestions at the time that *Jackson* had sounded the death knell for the negotiated plea. See, for example, the argument of the New Jersey Supreme Court in *State v Forcella*, 245 A (2d) 181 (N.J. 1968).

36 395 US 711 (1969).

37 Ibid. at 724.

38 Ibid. at 724, quoting *Worcester v Commissioner*, 370 F (2d) 713, 718 (1st Cir. 1966).

39 397 US 742 (1970).

40 Ibid. at 752.

41 Ibid. at 753.

42 Ibid. at 750.

43 Ibid. at 751 n.8.

44 Ibid. at 758.

45 Ibid. at 758. Adelstein (1978a) suggests that, apart from the effects of risk-aversion, the passage of time may play an important role in the decision to plead guilty. To the extent that this factor influences all defendants, there is a suggestion that innocent defendants may indeed be induced to plead guilty by the structure of the bargaining system itself.

46 See Griswold (1969, p. 314). English commentators have expressed a similar view with respect to the English criminal process. See, for example, Heberling (1978, p. 103).

47 'Of course, that the prevalence of guilty pleas is explainable does not necessarily validate those pleas or the system which produces them.' 397 US at 752–753.

48 404 US 257, 260 (1971).

49 *Scott v United States*, 419 F (2d) 264, 278 (DC Cir. 1969).

50 434 US 357 (1978).

51 395 US at 725.

52 434 US at 363.

53 Ibid at 364. In dissent, two Justices argued that this tactic was precisely the sort of 'vindictiveness' forbidden by *Pearce*. But the moment for these arguments was in *Brady*, not here; once the plea bargaining system has been vindicated, as the majority implies, it cannot be denied the single procedure that makes it work. One senses here that the real issue is submerged, for what makes Hayes' case so poignant and motivates the dissenters' passion is that his life sentence was triggered by the passing of a bad cheque for less than $90. This may indeed be a miscarriage of justice, but if it is, the real source is the harshness of the recidivist statute itself, an issue which was not raised in the case, rather than the bargaining tactics of the prosecutor.

54 [1968] 2 QB 108. See also *Davis* [1965] Crim LR 251; *Flynn* [1967] Crim LR 489.

55 [1968] 2 QB at 110.

56 Consider the discussion in *Scott v United States*, 419 F (2d) 264, 270–271 (DC Cir. 1969).

57 [1968] 2 QB 108.

58 Ibid. at 111. See also *House* [1978] Crim LR 173, decided in 1974, 1186/C/73, in which a sentence was reduced from seven to five years to reflect a guilty plea and 'for this reason only'.

59 See Baldwin and McConville (1978, pp. 117–118): '[T]he defendant who pleads guilty in the Crown Court is almost never asked by the judge if he wishes to say anything before sentence is passed.... There can be no doubt ... that if the courts had engaged in a more searching inquiry in the cases that we examined, they would have been rarely satisfied that the defendant was genuinely contrite.'

60 [1970] 2 QB 321.

61 Ibid. at 324.

62 Ibid. at 326.

63 Ibid. at 327.

64 Ibid. at 326–327.

65 The Times (London), 23 February 1976, at 11. Reported [1976] Crim LR 464.

66 The Times, ibid., at 11.

67 Ibid. (Widgery LCJ).

68 [1976] Crim LR 465.

69 'The decision in *R v Cain* has been subject to further consideration by the Court of Appeal. In so far as it is inconsistent with *R v Turner* the latter decision should prevail.' [1976] 1 WLR 799.

70 See, for example, *Grice* [1977] 66 Cr App R 167; *Llewellyn* (1978) 67 Cr App R 49; and similar cases discussed Baldwin and McConville (1979).

71 [1978] 2 All ER 460.

72 Ibid. at 461.

73 Ibid. at 462.

74 Ibid. at 461–462.

75 *Atkinson* [1978] 2 All ER 460, 463.

76 These points are discussed in detail in Adelstein (1978b, pp. 809–816).

77 [1977] Crim LR 47.

78 Ibid. at 48.

References

ADELSTEIN, R. (1978a), The plea bargain in theory: a behavioral model of the negotiated guilty plea, *Southern Economic Journal*, **44**, no. 3, 488–503.

ADELSTEIN, R. (1978b), The negotiated guilty plea: a framework for analysis, *New York University Law Review*, **53**, no. 4, 783–833.

ADELSTEIN, R. (1979a), Informational paradox and the pricing of crime: capital sentencing standards in economic perspective, *Journal of Criminal Law and Criminology*, **70**, no. 3, 281–298.

ADELSTEIN, R. (1979b), The moral costs of crime: prices, information, and organisation, in *The Costs of Crime*, (C.M. Gray, ed.) Sage, London, 233–255.

ADELSTEIN, R. (1981), Institutional function and evolution in the criminal process. *Northwestern University Law Review*, **76**, no. 1, 101–198.

BALDWIN, J. and McCONVILLE, M. (1977), *Negotiated Justice*, London: Martin Robertson.

BALDWIN, J. and McCONVILLE, M. (1978), The influence of the sentencing discount in inducing guilty pleas, in *Criminal Justice: Selected Readings*, pp. 116–128 (J. Baldwin and A. Bottomley, eds.), London: Martin Robertson.

BALDWIN, J. and McCONVILLE, M. (1979), Plea bargaining and the Court of Appeal, *British Journal of Law and Society*, **6**, no. 2, 200–218.

BLUMBERG, A. (1970), *Criminal Justice*, Chicago: Quadrangle.
BOTTOMLEY, A. (1973), *Decisions in the Penal Process*, London: Martin Robertson.
BOTTOMS, A. and McCLEAN, J. (1976), *Defendants in the Criminal Process*, London: Routledge and Kegan Paul.
CALABRESI, G. and MELAMED, A. (1972), Property rules, liability rules, and inalienability: one view of the cathedral, *Harvard Law Review*, **85**, April, 1089–1128.
COHEN, M. (1940), Moral aspects of the criminal law, *Yale Law Journal*, **49**, no. 6, 987–1026.
COX, S. (1976), Prosecutorial discretion: an overview, *American Criminal Law Review*, **13**, no. 3, 383–434.
CROSS, R. (1975), *The English Sentencing System*, 2nd ed., London: Butterworths.
GRISWOLD, E. (1969), Criminal procedure, 1969 — is it a means or an end?, *Maryland Law Review*, **29**, no. 4, 307–319.
HEBERLING, J. (1978), Plea negotiation in England, in *Criminal Justice: Selected Readings*, pp. 95–105, (J. Baldwin and A. Bottomley, eds.) London: Martin Robertson.
HOLMES, O. (1963), *The Common Law* (M. Howe, ed.) Boston: Little Brown.
JACKSON, R. (1972a), *The Machinery of Justice*, 6th ed., Cambridge University Press.
JACKSON, R. (1972b), *Enforcing the law*, rev. ed., London: Penguin.
LANGBEIN, J. (1977), *Comparative Criminal Procedure: Germany*, Minneapolis: West.
MICHELMAN, F. (1967), Property, utility, and fairness: comments on the ethical foundations of 'just compensation' law, *Harvard Law Review*, **80**, no. 6, 1165–1258.
MILLER, F. (1969), *Prosecution: The Decision to Charge a Suspect with a Crime*, Boston: Little Brown.
NEWMAN, D. (1966), *Conviction: The Determination of Guilt or Innocence Without Trial*, Boston: Little Brown.
PARKER, G. (1971), Copping a plea, *Justice of the Peace and Local Government Review*, **135**, 408–409.
POLLOCK, F. and MAITLAND, F. (1968), *The History of English Law, Volume II*, 2nd ed., Cambridge University Press.
POSNER, R. (1973), An economic approach to legal procedure, *Journal of Legal Studies*, **2**, no. 2, 399–458.
PRESIDENT'S COMMISSION on Law Enforcement and the Administration of Justice (1967), *Task Force Report: The Courts*, Washington: Government Printing Office.
THOMAS, D. (1979), *Principles of Sentencing*, 2nd ed., London: Heinemann.
THOMAS, P. (1978), Plea bargaining in England, *Journal of Criminal Law and Criminology*, **69**, no. 2, 170–178.
UNIVERSITY OF PENNSYLVANIA LAW REVIEW (1974), Just compensation and the assassin's bequest: a utilitarian approach, **122**, April, 1012–1032.
WILCOX, A. (1972), *The Decision to Prosecute*, London: Butterworths.
WILLIAMS, G. (1956), Discretion in prosecuting, *Criminal Law Review*, 222–231.