3 Criminal law and regulation

Anthony Ogus

1. Scope and focus of chapter

Few problems are encountered in delineating the scope of the criminal law since, as a matter of legal definition, it is the law prescribing liability for those offences which are prosecuted in the criminal justice system. Characterising particular forms of undesirable conduct as ‘criminal’ and subjecting them to the processes of the criminal law is thus one way of controlling such conduct.

‘Regulation’, in contrast, is not a legal term of art and its meaning varies according to the discipline of the discourse and the context (Mitnick, 1980, chapter 1). For the purposes of legal-economic analysis, it can be treated as having the following characteristics (Ogus, 2004a, pp. 2–3): it is public law, predominately enforced as well as formulated by an agency of the state; and it aims to induce outcomes which would not be reached by free market activity. It is therefore designed to overcome some perceived instance of market failure. As such, it can be usefully classified as ‘economic regulation’ where there is inadequate competition in the relevant market and ‘social regulation’, which is used to correct externalities and information asymmetry.

Now, of course, there are several regulatory methods of inducing individuals and firms to reach desired outcomes. Behaviour inconsistent with the outcomes may be made the subject of a tax or charge. Or regulatees may be required, as a matter of law, to refrain from such behaviour. The obligation may be incorporated as a condition for the lawful exercise of a licensed or authorised activity, with non-compliance leading to the suspension or revocation of that licence or authority. Breach of the obligation may, in the alternative, be the subject of a formal order by a court or other

---

1 I am grateful to Michael Faure, Niels Philipsen and the editor of this volume for comments on earlier drafts.

2 The governance of regulation is predominantly determined by administrative law, as to which there are major differences between legal cultures: most civil law jurisdictions have a coherent body of principles developed on the basis of what is required in the public interest and applied in a system of specialist tribunals; the common law approach involves the ordinary courts judging the legality of regulatory decision-making, primarily through judicial review. See Bell (2006).
public authority, with the threat of imprisonment, daily financial payments, or the seizure of assets in the event of continued non-compliance. Most obviously, the enforcement agencies may themselves have power to issue financial penalties, or to bring an action in the civil justice courts for such penalties. Last, but by no means least, the case can be taken to the criminal courts and the offender punished under the criminal law.

Across different jurisdictions and even within jurisdictions, there are widely differing approaches to the methods and practices of enforcing regulation, particularly by means of the criminal law. In some jurisdictions, criminal prosecutions are reserved for flagrant or very serious breaches while in others, for example, the United Kingdom, in theory at least they constitute the principal sanction for most regulatory contraventions; and indeed, there is evidence that enforcement agencies are increasingly relying on them (Baldwin, 2004). The latter approach has been criticised on the ground that it is unduly heavy-handed and unnecessarily expensive (Macrory and Woods, 2003; Ogus, 2004b; Hampton, 2005) and this has given rise to an important policy debate on the merits of using the criminal law for regulatory enforcement (Macrory, 2006). Clearly the policy question can be, and has been, addressed from an economic perspective and in this chapter I review the relevant literature.

To structure the review, I examine some of the key differences between the principles and procedures of criminal liability and the criminal justice process and their non-criminal equivalents. I begin (section 2) with the ex post character of most criminal offences, moving then to discussion of sanctions (section 3), followed by the conditions of liability (section 4) and procedural arrangements (section 5). I turn finally (section 6) to a consideration of the institutional characteristics of enforcement authorities and how they impact on efficient enforcement.

2. **Ex post deterrence versus ex ante prevention**

For the vast majority of criminal offences, liability arises only ex post engagement in the undesired, generally harm-creating, activity. The inducement to comply with the law thus rests on the assumption that the costs incurred by offenders as a result of the liability \( D \), when discounted by the probability of being caught \( p \), will be sufficient to deter them from engaging in the undesired activity because \( pD \) exceeds the benefit \( U \) to the offender of the offence (Becker, 1968).

In many cases, the costs arising from detected offences are significantly broader than the penalties imposed by the law since they include also the ‘hassle’ and personal inconvenience arising from encounters with enforcement officials, legal and other defence expenditures, as well as any loss of market reputation resulting from the detection. It follows that some of the
costs to the trader will be incurred even if there is no formal condemnation by a tribunal or other authorised institution. Because not all cases reach the ultimate stage of a court condemnation, involving the imposition of the formal sanction, it might be helpful to rewrite the ex ante additional cost arising from the contravention as $qE + pD$ where $qE$ represents the probability and the associated costs of the relevant act being detected, and $pD$ the probability of a formal condemnation and its associated costs, including notably the prescribed sanction (Ogus, 2006, pp. 102–5). The condition of compliance thus becomes $U < qE + pD$.

The theory of deterrence developed for the criminal law applies equally to many forms of regulation although, as we shall see in subsequent pages, differences between the regulatory and criminal processes have major implications for the efficiency of the deterrence. However, some regulatory forms operate to prevent, rather than deter, the undesired outcome.

The classic example is where an individual or firm must, prior to supplying a service or product, first obtain formal approval from some public agency, and to secure the approval the applicant must show that the product or service satisfies certain standards (Ogus, 2004a, chapter 10). The idea may be that the screening process should ensure that undesired outcomes do not occur, but the economic arguments for prevention, rather than deterrence, are not always compelling and must be carefully considered (Moore, 1961; Shavell, 1984 and 1993). If the benefits of particular ex ante regimes do exceed their costs, that may provide a public interest justification for their existence, as an alternative to the more familiar private interest, rent-seeking explanation based on public choice analysis (Svorny, 2000).

A prima-facie case for ex ante prevention can be made out for situations where ex post deterrence is likely to be ineffective. One possibility is that the sanction and other associated costs of ex post liability might be insufficiently high relative to the benefit which the offender acquires from committing the crime. Although the financial level of the penalty might be increased, that could exceed the offender’s assets and hence have little impact. Another possibility is inadequate enforcement, and hence a low value of $p$.

In principle, both of these problems can be overcome, the first by substituting a non-financial penalty, such as imprisonment, and the second by allocating additional resources to monitoring behaviour. However, both policies are costly and the task is, then, to compare the benefit-cost ratio (the amount of social harm avoided from deterring the undesired outcome, relative to the administrative costs incurred) of a deterrence policy with the equivalent ratio for a regulatory preventative policy, which as we shall see, gives rise to large administrative costs.
For the purpose of this comparison, the amount of harm caused by the undesired activity may have relevance. The larger the amount, the more likely it is that risk-averse politicians, concerned with public reactions to the risk of very large losses, will opt for a regulatory ex ante preventative policy, thus suggesting that, in this context, some additional weighting must be attributed to the benefit of preventing such losses (Posner, 2004, p. 199).

Another variable relevant to the comparison and discussed in the literature (Malik, 1990; Nussim and Tabbach, 2007) is that of detection avoidance. Under an ex post deterrence regime, potential offenders are motivated to use resources in avoiding detection and thus punishment. This is not only wasteful, since no social benefits accrue from the behaviour; it also impedes deterrence. Under an ex ante prevention regime, the opportunities for avoidance are much more limited.

The prevention of harm is, in principle, secured in ex ante licence regimes by the systematic scrutiny of all market participants. Of course, such a system generates very high administrative costs, compared with a deterrence regime, under which the enforcement agency is concerned only to investigate cases of suspected contravention and therefore will monitor the behaviour of only a proportion of the market participants. The principal trade-off to be considered by policy-makers thus appears to be between the higher social benefits of prevention at higher administrative costs and the lower social benefits of deterrence at lower administrative costs. In fact, the comparison is significantly more complex.

Take, first, the administrative costs of enforcement. Although deterrence systems involve less monitoring, those contraventions which are detected will give rise to the additional costs of prosecuting offenders and imposing sanctions on them. Under an ex ante regime, although these costs will not arise in relation to those deemed by the screening process to be unfit for a licence, monitoring of the market activity has to take place to ensure that unlicensed suppliers are not participating (Ogus and Zhang, 2005). Furthermore, since licensing excludes some suppliers from the market, the process can be exploited by private interests to create artificial barriers to entry, thus reducing competition and generating welfare losses (Maurizi, 1974; Gellhorn, 1976): there is much empirical evidence of anti-competitive practices being linked to licensing regimes (surveyed

---

3 As Nussim and Tabbach (2007) point out, it may also lead to the paradox that a more severe punishment might induce more crime, because potential offenders will make greater efforts to avoid detection.

4 The cost of such monitoring can be further reduced by risk assessment and profiling.
in Svorny, 2000). Since the satisfying of conditions for a licence generally calls for some discretionary decision-making, this confers power on officials and in some circumstances can give rise to bribery and other corrupt transactions (Manion, 1996).

In short, given the heavy administrative and welfare losses which are consequent on most ex ante preventive regimes, the onus of proof must be on those who contend that, in relation to particular areas of regulation, the benefits exceed those costs. Such may be the case where ex post deterrence is likely to prove to be ineffective or where very high social losses can result from a single contravention and, for this latter purpose, ‘social losses’ should be construed to include fear generated by the regulated activities.

3. Sanctions

A major difference between criminal and non-criminal methods of enforcing regulatory obligations arises from the sanctions, both formal and informal, which are available. The penalties which a criminal court can impose on conviction for an offence include imprisonment and community service as well as fines and the confiscation of assets. In the absence of a criminal prosecution, the administrative authorities may, exceptionally, have powers to revoke or suspend a licence, or to obtain from a court an injunction, or equivalent order prohibiting further non-compliance, breach of which may lead to proceedings for contempt or a per diem financial charge. In most, but not all, jurisdictions, financial penalties may be imposed by a civil court or (generally to only a modest level) by the administrative authorities themselves. Other sanctions include the issuing of adverse publicity (‘naming and shaming’), and the seizure or immobilisation (for example, ‘wheel-clamping’) of assets.

From a deterrence perspective we need to analyse the impact of these different regimes and to hypothesise on the circumstances in which use of the criminal justice sanctions might be optimal. We note first that the cost to the offender following conviction in the criminal process will generally be significantly higher than that which follows non-criminal procedures. This is not only because the formal sanctions available in the criminal process are, in broad terms, more severe – notably with reference to imprisonment and very large fines; it is also because stigma attaches to a criminal conviction much more than to other forms of condemnation

---

5 The administrative power to revoke or suspend a licence may be an exception.
(Funk, 2004). Indeed, there is empirical evidence that loss of market reputation following criminal proceedings may be of greater consequence to some firms than the penalty imposed (Baldwin, 2004, pp. 361–4; Karpoff et al., 2005).

Now if, for the purpose of Becker’s model of deterrence, the criminal process can be used to ensure that the offender, on conviction, will face costs \( pD + qE \) sufficiently high to exceed the utility \( U \) to be derived from the contravention, why should it not be invoked in all cases? There are four important answers to that question and it is a consideration of these which should lead us to the means of identifying when criminal sanctions might be optimal for regulatory contraventions.

Two arguments relate to the lowering of costs which are generated by the imposition of criminal penalties. The first has regard to the costs incurred by the offender. Although it is of the very essence of a deterrence system that individuals and firms should incur costs if they offend, social welfare is increased if the same level of deterrence can be achieved when those costs are smaller, rather than larger. Secondly, account has to be taken of administrative costs. Of course, some criminal penalties, notably imprisonment, are much more expensive to apply than administrative or civil penalties; but more significant is the comparison of process costs, since (as we shall see) the procedural requirements of criminal liability are much more exacting than their administrative and civil equivalents. In short, the goal should be the minimisation of costs, consistent with the securing of adequate deterrence.

The third consideration is that of marginal deterrence (Friedman and Sjostrom, 1993; Mookherjee and Png, 1994): other things being equal, the severity of penalties should increase proportionately with the severity of the harm caused by illegal activities. This is because offenders intent on committing a less serious offence should not be motivated, perversely, to commit a more serious offence for the same penalty. If a severe criminal penalty is handed out for a relatively minor regulatory contravention, then more severe penalties, at additional administrative costs, will have to be handed out for more serious contraventions.

Fourthly, excessively severe penalties can lead to over-deterrence (Bierschbach and Stein, 2005). It may be the case that with mainstream crimes, such as theft, assault or arson, there is little or no social utility in the behaviour which constitutes the offence, but regulatory offences

---

6 The economic interpretation is that since procedural safeguards reduce the number of erroneous decisions, a criminal conviction conveys more reliable information than decisions from other institutions: Galbiati and Garoupa (2007).
are different (Selznick, 1985, p. 363). The latter normally arise during the course of everyday activities, including industrial and commercial undertakings, which enhance social welfare. Of course, with insufficient precautions, those activities may generate harm which exceeds the social benefit, hence the regulatory control, but if the prospect of a severe penalty induces the firm either to reduce the amount of the activity or to invest in an excessive level of care, or both, social welfare losses are incurred.

If the discussion above indicates why criminal penalties should not be used for all regulatory contraventions, there are other powerful arguments suggesting that administrative sanctions are preferable for most such contraventions. Take, first, the question of the offender’s knowledge and information concerning legal obligations (Craswell and Calfee, 1986). This is generally not a problem in relation to mainstream crime, covering the familiar range of offences to the person and property, because, as the Latin expression (\textit{mala in se}) used to distinguish them from regulatory offences (\textit{mala prohibita}) indicates, these are inherently wrongful (Wiley, 1998): all individuals, except those who are mentally disabled, are presumed to have a sufficient understanding of the law. Most regulatory contraventions, in contrast, involve complex obligations and many regulatees, particularly if they are individuals or small firms, in the absence of some interaction with regulatory agencies, have only a sketchy knowledge of what is required (for empirical evidence, see, for example, Frick and Walters, 1998; Environment Agency 2005). Where a regulatory contravention occurs through inadvertence or ignorance as to legal requirements, it is inefficient to secure future compliance by the costly imposition of a criminal penalty if the communication of appropriate information by the regulatory agency will be equally effective in securing this outcome (May and Wood, 2003).\footnote{Although the strategy imposes no formal sanction leading to costs \(D\), it should not be forgotten that mere detection can impose large informal costs \((E)\) on the offender and ex ante \(qE\) may not be insignificant relative to \(U\).}

It is, of course, far too easy for an offender to plead inadvertence or ignorance and the cost to the authorities in securing the evidence to rebut the claim will be disproportionately high. As we shall see, that is a justification for applying strict liability in the application of administrative sanctions, but the consideration can also lead us to a further generalisation, distinguishing between first offenders and second (and subsequent) offenders (Ogus et al., 2006). If the empirical evidence referred to in the last paragraph is reliable, it is not likely that imposing a penalty on first offenders will ensure future compliance better than dismissing the case with advice and/or an informal warning (Scholz and Gray, 1997; Nyborg
and Telle, 2004). Note, too, that a heavy-handed approach, for little or no additional benefit, will not only generate increased administrative costs; it will also involve significant additional costs to regulatees. And these are likely to include ‘indignation costs’ which may motivate them to initiate appeals against decisions, leading to yet more administrative costs (Bardach and Kagan, 1982).

The argument does not apply where a regulatory contravention gives rise to large losses: here even a first offence should be deterred. But subject to that important caveat, the reasoning leads to the conclusion that generally, in a regulatory context, substantial, particularly criminal, penalties should not be imposed for a first offence. That reasoning helps to explain the empirical finding that, in most jurisdictions, in the absence of deliberate wrongdoing or serious harm, regulatory agencies tend to be lenient towards first offenders and the insight has formed the basis of the theoretical analysis in the literature.

Fenn and Veljanovski (1988) reason that, since the monitoring costs of enforcement agencies are high, it is in their interest to secure a commitment to future compliance by an agreement not to prosecute a past offence. Of course, for the offender to be motivated to adhere to the agreement, there must still be a credible threat to prosecute on the occasion of the second bite of the cherry. But since the firm in question will be aware that it is now ‘under observation’, the probability of being apprehended is significantly increased – compared to its first contravention. And hence a higher perceived value to the \( pD \) of a formal conviction might be plausible.

To similar effect, Harrington (1988) has modelled the relationship between an enforcement agency and firms as a game in which profit-maximising firms have to decide whether or not to comply with the law, and the agency, aiming to maximise compliance, has to decide whether or not to punish a firm which decides not to comply (see also McCannon, 2006). For the purposes of the game, the agency classifies all firms into two groups, \( G_1 \) and \( G_2 \): the first is monitored only rarely, while the second is monitored regularly and, if found to contravene, receives a substantial penalty. Initially, the agency does not have sufficient information on which to classify firms. The first monitoring enables it to do so. If a firm is found to comply, it is classified as \( G_1 \).\(^8\) If it is found to have committed a contravention, it is classified as \( G_2 \); it will not be punished, but it will know that another contravention will lead to the prescribed punishment and the

\(^8\) The game can be played opportunistically by a firm, ensuring that it is seen to comply on the first inspection, and only then deciding to contravene. Hence the importance of some residual monitoring.
chance of the contravention being detected is now relatively high, given that it will be subjected to regular monitoring. So it must be seen regularly to comply, thus enabling it to be reclassified as $G_i$. Institutional arrangements which facilitate the game being played to these outcomes should lead to the goals of firms and agencies being met at lowest cost.

Both of these models can be accommodated within the well-known ‘enforcement pyramid’ of Ayres and Braithwaite (1992, p. 35). The horizontal axis of the pyramid represents the percentage of apprehended contraventions for a given offence, and the vertical axis the steps of increasing severity which can be taken in the treatment of the offender from, at the base, an informal warning to, at the apex, a substantial criminal penalty. In accordance with the economic models, the marginal benefits of the enforcement agency proceeding up the pyramid until it is satisfied that it has secured a particular offender’s compliance should exceed the marginal costs.

Reference to this pyramid also reveals the problem of excessive reliance on criminal, as opposed to administrative, sanctions in deterring regulatory contraventions (Woods and Macrory, 2003; Ogus, 2004b). Although the costs ($D$) to the offenders increase with each higher stage, so also will there be a significant reduction in the ex ante probabilities, $p$ and $q$, that the relevant next stage will be reached. In relation to the criminal sanction at the apex of the pyramid, $p$ is likely to be small, particularly because of the disproportionately high cost to the agency of procuring evidence sufficient to satisfy the procedural standards of liability in the criminal courts (see further on this below). If criminal liability constitutes the only significant penalty, then, for deterrence purposes, this must be sufficiently large when discounted by $p$ to exceed the profit to the offender from the contravention. Since courts are typically reluctant to impose large penalties for most regulatory contraventions (Macrory, 2006, pp. 30–31), such a system is likely to suffer from under-deterrence. Clearly this problem should not arise if the agency is able to impose an administrative financial penalty at an intermediate point in the pyramid, where $p$ will have a relatively high value (for empirical evidence, see Brown, 1992).

The economic models formulated for the purpose of determining the optimal criminal fine and incorporating variables of the offender’s wealth and the probability of apprehension and conviction (for example, Polinsky and Shavell, 2002) are obviously applicable to administrative financial penalties, but there are some special considerations arising from

---

9 In Europe, $p$ rarely exceeds 0.05: see the empirical studies cited in Ogus and Abbot (2002, pp. 497–8).
variations in the mode of exacting such penalties. There is a spectrum of possibilities ranging from a financial charge automatically imposed for each contravention and a discretionary, but fixed, penalty payable to an officer of the agency when a contravention is detected, to a discretionary variable penalty selected after a reviewing of the circumstances of the case (Macrory, 2006, pp. 41–52).

Clearly, as the selection of sanction moves across the spectrum from fixed and automatic to variable and discretionary penalties, so there is increased deterrence targeting, but at greater administrative cost. The trade-off between these benefits and costs will vary according to the volume of cases, the amount of social harm resulting from contraventions, the level of information of regulatory requirements to be expected of offenders, and the degree of variation in their wealth. The greater the social harm relative to the number of contraventions, the greater the degree of wealth variation and the lower the information level of regulators, the greater the benefit from targeting, thereby justifying the higher administrative cost. Account has also to be taken of the fact that the larger the penalty imposed, the higher the costs of erroneous decisions, and therefore the greater the benefit from more elaborate procedures (see further on this section 5, below).

Mention can also be made here of the deterrence effectiveness of regulatory sanctions which have little or no equivalent within the criminal process. We have already seen that the latter generates stigma which may be important in raising the costs above those arising from the formal penalty to a level necessary for deterrence purposes. While an administrative imposition will not by itself have this effect, a similar outcome can be achieved by the regulatory agency publicising the fact in some appropriate media (Whitman, 1998). Likewise, the confiscation or immobilisation of assets can facilitate deterrence where the limited wealth of offenders renders them insensitive to financial penalties but where the costs arising from draconian penalties such as imprisonment are not justified by the level of harm inflicted by the contravention (Bowles et al., 2000 and 2005). Empirical evidence suggests that, in this respect, the wheel-clamping of private vehicles is a particularly effective sanction (Halcrow, 2007, Appendix D2).

None of this is to deny the importance of criminal liability as a residual deterrent for certain categories of regulatory offences and regulatory offenders, notably where a single contravention gives rise to a significantly large amount of harm and for repeated contraventions when other sanctions provide inadequate deterrence. As Garoupa and Gomez-Pomar (2004) demonstrate, resort to a criminal sanction may be economically justifiable also where the institutional characteristics of the enforcement
agency, for example, the existence of conflicts of interests or political influence, result in a less than optimal level of deterrence; so also where the regulatory penalty is larger than the reward received by the agency and where, therefore, there is a possibility of collusion between the agency and offenders.

4. Conditions of liability: mens rea

Prosecutions for criminal offences are subject to very different liability regimes and procedural arrangements than those giving rise to administrative sanctions. Most obviously, criminal liability normally requires proof of intention or at least some form of blameworthiness (mens rea), while these conditions are generally absent from regimes governing administrative sanctions. What explanations are there for this different approach?

The traditional legal rationalisation of mens rea relates to the moral basis of the criminal law (Ashworth, 2000): since criminal liability implies moral obloquy, it is necessary to ensure that the offenders assume some degree of mental responsibility for their acts. In their (perhaps unnecessary) strenuous efforts to avoid this simple explanation, exponents of law and economics have proffered several economic arguments, none of which is entirely convincing.

Perhaps the best known are those of Posner (1985) who articulates three economic rationalisations for mens rea. The first arises from his defining crime as a ‘pure coercive transfer’, which deliberately avoids voluntary market transactions that could have been accomplished at relatively low transactions costs. Since there is little or no social utility in the act, the amount of the offender’s gain is irrelevant and all crimes therefore should be deterred. This is contrasted with what Posner calls ‘impure coercive transfers’, which includes what in the context of this chapter are regulatory offences. Although these acts to some extent bypass the market, that aspect is ancillary to the main purpose which has social utility. In the ‘pure’ case, the actor intentionally uses resources to inflict the harm which avoids the market transaction; in the impure case, most of the resources are deployed for market-related purposes and the harm is only incidental. Now, although this analysis might be helpful in identifying mainstream crime with the absence of social utility, it provides an imperfect justification for the mens rea requirement as used in the criminal law, because the latter covers a far wider range of mental states (for example, acting with the knowledge that the harm is highly likely to result\(^\text{10}\)) than deliberate

---

\(^{10}\) See, for example, in English law *R v. Woollin* [1999] 1 AC 82.
or intentional harm, which is the central feature of Posner’s argument (Parker, 1993; Finkelstein, 2000).

The other two economic rationalisations are linked to the goal of optimal enforcement. An intended harm is more likely to occur, more likely to be concealed and (perhaps) more likely to lead to very substantial losses, compared with unintended harm; and therefore there is a strong argument for concentrating resources on deterring such activity. These contentions are subject to the same objection, that they identify *mens rea* too closely with intention. But quite apart from that, they lead only to the proposition that intentional wrongdoers should face a heavier penalty than unintentional wrongdoers. It does not follow that that penalty should necessarily be imposed by the criminal law.

Very different economic rationalisation of *mens rea* are offered by Parker (1993) and Hamdani (2007), focusing on the information costs to the offender of responding to legal obligations. Parker, having observed that the prospect of high penalties will induce potential offenders to invest resources in acquiring information relevant to criminal liability, argues that such investment will be economically excessive where the penalty is high relative to the social harm caused by the act. In such circumstances, the requirement of *mens rea* discourages the investment and is therefore economically justified. Normatively, the reasoning may convince, but it is less easy to use it as a rationalisation for the difference between mainstream crime and regulatory offences, since most mainstream crimes seem not to fit the pattern of a relatively high penalty combined with relatively low social harm.

Hamdani’s approach provides, perhaps, a better explanation. As he notes, strict liability, that is liability without *mens rea* typically adopted in administrative processes, induces potential offenders to acquire information relevant to liability which would not readily be available at low cost to them as a consequence of ordinary market processes. The requirement of *mens rea*, in contrast, creates an incentive to remain ignorant in such circumstances. As a broad generalisation, it is probably the case that information relevant to mainstream offences is typically available at low cost, while this is not the case for many regulatory offences given their complexity and technical content.

5. **Procedural arrangements**

A criminal penalty can be imposed only by a court – this is a requirement of human rights law (Guinchard, 2005) – and only following completion of prescribed procedures. These vary between jurisdictions and particularly across legal cultures, but the procedures generally include: (for more serious offences) trial by jury; a burden of proof on the prosecution to a
standard higher than that used in non-criminal cases; and rules of evidence protecting the defendant. In contrast, some administrative sanctions, for example, small fines and financial charges, can be imposed without an oral hearing or other formality, although an alleged offender, unwilling to make the payment, might be provided with the opportunity to offer reasons why the sanction is inappropriate, perhaps by way of appeal to a court or tribunal. Even with this last qualification, it is clear that there are very major procedural differences between criminal and administrative processes: what are the economic justifications?

The principal purpose of the procedural arrangements is to reduce the number of errors that can result from the adjudication process, including notably the wrongful imposition of a penalty on those who are innocent, and the costs that arise from such imposition. But since the procedures designed to reduce errors are themselves costly, there is in principle an optimal level of error reduction where the benefits of such reduction approximate to their costs (Posner, 1973; Lando, 2003). For the differences between criminal and administrative procedures to be economically justified, we therefore need to identify the relevant variables and variations in the values to be attributed to them.

Take, first, the benefits to be derived from reducing wrongful penalty impositions, sometimes referred to as Type I errors. Most obviously, the innocent parties themselves benefit from the procedural arrangements since the penalties and other losses associated with them are avoided ($B_1$) and it seems appropriate to add a set of less concrete benefits ($B_2$), in that society gains from increased confidence in the quality of justice generated by its institutional processes (Polinsky and Shavell, 2000).

To be set against these benefits are the costs. These may include any increase in administrative resources which have to be spent on the adjudicative process ($C_1$). Further costs ($C_2$) arise from the impairment of the deterrence function of the law, resulting from the fact that some guilty defendants will avoid the penalty (Type II errors). It is important to appreciate that the reduced deterrence resulting in $C_2$ occurs for two related but independent reasons. Some of the ‘guilty’ cases processed will not satisfy the higher procedural standards and will fail for that reason. Others will not be prosecuted because, with limited resources, the enforcement agency will concentrate those resources on clear-cut cases requiring less prepara-

\footnote{Posner (1973, p. 412) and Png (1986) rightly argue that reducing Type I errors should enhance deterrence because it will increase the net punishment cost of committing the crime relative to that of abstaining from the crime, but this effect will surely be outweighed by the considerations mentioned in the text.}
tion (Polinsky and Shavell, 1989; Miceli, 1991 and for empirical evidence, Ogus et al., 2006).

To compare the incidence of these benefits and costs in relation to criminal and administrative procedures, take, first, by way of example, the question of burden of proof, which in criminal law is on the prosecution but which in many administrative systems is, in practice at least, on the suspected offender. A simple guideline here is to place the burden on the party who can supply the information at lower cost, and it may be that, in this respect, the regulatee-regulator relationship has different characteristics from the criminal defendant-prosecutor relationship (Lewis and Poitevin, 1997). But there are other variables involved. The criminal law approach operates to avoid substantial adjudication and post-adjudication administrative costs, including the imposition of penalties, in cases where liability is indeterminate: placing the burden on the defendant would therefore significantly increase \( C_1 \). It also reduces Type I error costs \( (B_1) \) at the expense of Type II error costs \( (C_2) \) (Lee, 1997).

In the context of the criminal justice system, the burden of proof rule is economically appropriate because adjudication and post-adjudication administrative costs are particularly heavy, and because, given the stigma attaching to criminal law convictions, Type I error costs are assumed to be significantly higher than Type II error costs: it is ‘better that ten guilty persons escape, than one innocent suffer’ (Blackstone, 1836, p. 358). Adjudication and post-adjudication administrative costs are generally less onerous in a regulatory non-criminal context, but the principal economic justification for the alternative burden of proof rule must be that, given typically relatively modest penalties and an absence of stigma, Type I error costs are not so significant and, indeed, may be outweighed by Type II error costs (Posner, 1973, pp. 414–16).

The analysis has similar implications for other procedural protections conferred on criminal defendants, including notably the standard of proof which is particularly important, since there is a striking contrast between the criminal rule (‘beyond all reasonable doubt’) implying a level of accuracy above 0.95 and the equivalent administrative standard, which is just above 0.5 (Kokott, 1998, p. 134).12 Other relevant rules include those restricting the evidence on which the prosecution can rely (on which see Schrag and Scotchmer, 1994; and for a survey of the relevant literature, Parker and Kobayashi, 2000). A quite different economic approach to

---

12 The text reflects the rules in common law jurisdictions; civil law jurisdictions have a different approach. For an economic analysis of the comparison see Demougin and Fluet (2005).
these phenomena is offered by Hylton and Khanna (2007). They argue that while the analysis above explains the pro-defendant bias in American criminal procedure, it does not justify the extent of the protections and the latter are better explained by reference to public choice considerations. From this perspective, since the protections increase the costs to self-interested actors of using the criminal process for their own ends, they help to reduce the rent-seeking and deterrence costs associated with abuses of prosecutorial authority.

6. Institutional considerations

In terms both of rule-making and of enforcement, there are many institutional differences between regulatory processes and criminal justice processes, but we need only highlight those which are the most significant for the purposes of this chapter. Undoubtedly, the major such difference relates to the separation of powers. The rules governing criminal liability are, in theory at least, determined by the legislature; compliance with the law is principally monitored by a police force; they may undertake prosecutions, but in many jurisdictions this is the responsibility of an independent authority; and liability is determined by a court (often by a jury) and the appropriate penalty is selected by the judiciary. In contrast, in relation to regulatory controls, although the legislature establishes the principles, the detailed rules are typically formulated by a regulatory agency (which may be more or less independent of government); and that same agency is normally responsible for monitoring behaviour and initiating administrative enforcement proceedings. These roles may, in some areas and to some extent, be shared by the regulated industry under co-regulation or self-regulation arrangements. In some jurisdictions, liability for, and selection of, the administrative sanction is the responsibility of an independent authority or tribunal, but in others they are undertaken by the regulatory agency,13 with the offender having access to an independent authority only by way of appeal.

Generalising on these arrangements, it may be said that regulatory institutions and functions (rule-making, monitoring, enforcement and adjudication) are, to a large degree, vertically integrated, whereas in the criminal sphere they are separated. The integration generates considerable savings in the coordination and information costs of regulatory processes, but of course it also has adverse consequences. An independent assessment of the

13 The character of agencies differs across jurisdictions and legal cultures, as a consequence of divergent approaches to administrative law: see above, n. 2. The civil law tradition is for agencies to be part of, or close to, central government; the common law tradition is of greater autonomy: Ogus (2002).
application of the rules can reduce intertemporal inconsistency and hence error costs (Stephenson, 2006). Further, some form of competition of, or overlapping between, decision-makers can inhibit corrupt transactions (Rose-Ackerman, 1978; Bowles, 2000).

For an economic appraisal of these arrangements, we can begin with rule-making. Mainstream crime tends to be governed by general principles rather than specific rules. On the assumption that they will be properly interpreted to meet the circumstances of individual cases, general principles avoid the over-deterrence and under-deterrence arising from specific rules, but the superior targeting comes at the price of increased administrative costs (Ehrlich and Posner, 1974; see also Stephenson, 2006, who models the legislative choice in terms of diversifying risk across time and issues). Given the higher error costs, particularly of Type I, associated with criminal justice adjudication, it is appropriate to reduce these by general principles, albeit at an increased administrative cost.

As we have seen, the incidence of administrative and error costs in a regulatory context varies significantly and specific rules with much technical detail therefore predominate. Delegation to the regulatory agency for fleshing out the skeletal legislative principles can be justified in terms of economising both on information costs and on the costs of rule promulgation, given the likelihood of frequent amendment (Mashaw, 1985). Of course, this gives rise to a principal-agent problem (Schwartz et al., 1994), but judicial review and other administrative law devices can contain this (Bishop, 1990; and for a survey of the substantial literature on this question, see Spiller, 1998). The delegation of rule-making to regulatory agencies can also be explained on public choice grounds (Fiorina, 1982; Aranson et al., 1983): politicians can benefit from a generalised legislative response to the competing demands of pressure groups; the burden of the costs arising from disappointment at the outcome of implementation can be shifted to those responsible for drafting the detailed rules.

The public interest arguments for delegation to agencies can be extended to justify co-regulation and self-regulation (Tuohy and Wolfson, 1978): the regulated industry generally has superior technical knowledge and expertise to that of agencies; and there are even greater economies in the costs of rule promulgation. This is because it is cheaper to publish and disseminate the rules, which may be only informal, within the private community. At the same time, and obviously, the principal-agent problem

---

14 The deterrence effect of uncertain standards is analysed in Craswell and Calfee (1986). They conclude that where the uncertainty is small, over-compliance is likely to occur, but broad uncertainty is likely to lead to under-compliance.
becomes acute with such arrangements, given the motivation of the regu-
lated industries to exploit their (partial) rule-making and perhaps enforce-
ment powers to generate rents for themselves (Shaked and Sutton, 1981).
To meet it requires effective accountability to a public agency (‘enforced self-regulation’: Ayres and Braithwaite, 1992) or competition between dif-
f erent self-regulatory agencies (Ogus, 1995).

The borderline between rule-making and enforcement is not always easy
to define because the more general the regulatory requirement, the greater
the agency’s discretion and the more important interpretation becomes in
the enforcement process. We have also seen, in these pages, that discre-
tion in deciding whether or not to take formal proceedings in relation to
contraventions plays an equally significant role, because supplying infor-
mation, particularly to first offenders, can provide a less costly means than
deterrence of securing compliance.

Indeed, explicit, or more usually implicit, bargaining between regula-
tor and regulatee can help to generate a wide range of efficient outcomes
(Yeung, 2004). In analogy to plea-bargaining, agencies may agree to
informal standards, lower than those required by the law, in return for a
commitment to compliance (Kambhu, 1989; see also Johnston, 2001, who
indicates the importance of the device in relation to environmental regu-
lation where there may be information problems and some uncertainty
regarding future requirements). Law (2006), drawing on an empirical
study, shows how enforcement agencies, with a limited budget, can gain
the cooperation of regulatees by offering benefits other than a decision not
to prosecute for a contravention. The agency may have technical expertise
not available to small firm regulatees and can assist the latter by supply-
ing information as to low-cost methods not only of meeting regulatory
requirements but also of attaining higher quality output more generally.

Of course, the phenomenon of arm’s-length bargaining between offender
and enforcement agency raises the possibility of private agreement which
serves to benefit both parties, but not the public interest. In the absence
of adequate controls involving transparency and accountability, this can
lead to corrupt transactions with major negative externalities (Polinsky
and Shavell, 2001; Garoupa and Klerman, 2004), as well as high costs to
business and incentives to move towards the unofficial economy (Johnson
et al., 1998). These problems are so prevalent in less developed countries
that the costs of discretion within the regulatory institutional arrange-
ments may outweigh the benefits (Ogus, 2004c). An independent argument
for rules rather than discretion in less developed countries is the human
capital argument that discretion generates demand for a better educated
workforce, which might be in limited supply (Schaefer, 2002).

A final issue concerns the role of enforcement agencies in formally
determining that a regulatory contravention has occurred and selecting and imposing the administrative sanction. To be sure, there is here an overlap with the characteristics of financial penalty, considered above: an ‘on the spot’ fine can hardly be administered otherwise than by the agency responsible for detecting the contravention. In other respects, the issue turns on the familiar trade-off between error costs and administrative costs. There are, of course, large savings of the latter if the same institution is responsible for different stages in the enforcement process. But such arrangements can also generate large error costs, and decision-makers can be prone to commit such errors if the agency itself benefits from the volume of formal condemnations recorded for contraventions (Posner, 1973, pp. 416–17). Such benefits may arise in two forms: an enhancement of the public reputation of the agency for effective performance of its enforcement responsibilities; and additions to the agency’s income if financial payments accrue directly to it. The normative implications are clear: although there is little that can be done to moderate the first benefit, the second should not be permitted; and, in any event, the offender must be able to initiate some form of appeal against, or review of, the decision. The optimal trade-off between administrative and error costs is dependent on several important variables. These include notably: the probability of error which itself is a function of the procedural arrangements, in particular the facilities for appeals and the size of the penalties imposed.

References


Criminal law and economics


