

Published as: ARLC Conference Papers

Penalties: Policy, Principles and Practice in Government Regulation

June 2001

Managing Discretion

Dr Julia Black

Law Department and Centre for the Analysis of Risk and Regulation
London School of Economics
Houghton Street
London
WC2A 2AE
UK
Tel +44 (0)20 7955 7684 (sec)
Email j.black@lse.ac.uk

Please note this paper is available on a CD from the ALRC.

Introduction: Understanding Discretion

Having discretion means having the freedom to choose between courses of action. Giving someone discretion is giving them that freedom. If there is no concern as to how that freedom is used, then there is little to detain us. But more often we are concerned that the choices that are made should be the 'right' ones in some sense. We rely on the person who has discretion to act in an appropriate manner. Discretion is then in this context intimately related to trust: the more the person is trusted to make the right choices as to courses of action, the more it is felt discretion can be given to them to make those choices. The less they are trusted, the more it is sought to reduce that discretion to confine those choices in some way.

Concerns that bureaucratic discretion may lead to arbitrariness, inconsistencies, unpredictability and decisions based on officials' personal predilections have been commonplace ever since the displacement of the Weberian bureaucrat as the model of bureaucratic behaviour. Correspondingly, the call has traditionally been to limit and confine that discretion. The assumptions have been, as Davis argued, that discretion can be 'managed', it can be controlled, directed, limited, structured, and that the way to do that is through rules. Discretion, Davis famously argued, can be 'structured' and 'confined' through rules, and can be 'checked' through mechanisms of accountability (Davis 1969; see also Lowi 1969). The more rules there are, and the more precise they are, then the more the freedom to choose between courses of action can be curtailed.

But as many have argued since, discretion and rules are not in a zero sum relationship such that the more rules there are, the less discretion there is and vice versa. Discretion is something that can be present in degrees, it is true - one can have more or less discretion, be more or less bound (Dworkin 1977: 31-39; 68-71; for criticism see Galligan 1986: 14-20). But it is always present in both more and less degrees than one might think, and is exercised in places that one might not expect. Far from being easily confinable or structurable by rules, rules and discretion operate in a complex interrelationship (Baldwin and Hawkins, 1984, Galligan 1986; Hawkins 1992b, Bell 1992, Schneider 1992; J. Black 1997). Further, far from discretion being easily distinguishable from rules, as in Dworkin's imagery of discretion as the 'hole in the doughnut' (Dworkin: 1977), the very nature of rules, their need for interpretation in application, means that rules are shot through with discretion. Discretion is the space both within and between rules in which legal actors exercise choice (J. Black 1997: 216; cf. Hawkins 1992b; Galligan 1986). Moreover, how decision makers make decisions is only partly determined by rules, be they organisational or legal rules. Organisational norms and practices, past experiences, personal relationships, the decision maker's own perceptions and attitudes will all play a part in affecting how decisions are made. Thus the presence of rules does not mean that rules will be the sole or even dominant factor influencing how discretion is exercised, and their absence does not mean the decision maker is unbound in his or her decision: bureaucratic and organisational norms will continue to operate, as will broader political and economic pressures, and moral and social norms. (Hawkins, 1992a, 1992b, 1986; Hawkins and Manning 1990; Galligan 1986; J. Black 1997b).

The assumption, therefore, that discretion can be 'managed' both to confine its exercise to certain actors and to limit the way that those actors use the discretion given, is thus ambitious, if not misguided. Further, the idea that some legal actors can be completely deprived of discretion is simply erroneous. Part of the remit I was given for this paper asks who should exercise discretion: administrative officials, courts or tribunals. The answer is that the question is irrelevant. All those involved in the regulatory and legal process will be exercising discretion. As Hawkins has noted, '[d]iscretion is a central and inevitable part of the legal order' (1992b: 11). It is central because of the broad grants of authority to legal and administrative officials to attain broad legislative purposes

which is the hallmark of much current regulatory law. It is inevitable because the application of rules, their translation into action, involves interpretation and choice (ibid; see also Galligan 1986: 1). We can debate who should have the authority to make final decisions as to the rule's application (Dworkin's second form of weak discretion: (Dworkin 1977)) but we cannot assume that we can confine the exercise of discretion to only certain types of decision maker (courts, tribunals, agencies) or even to particular levels within an organisational hierarchy. Every person at every level who is involved in considering whether particular practices should be seen to be in conformity with the law and what should happen if they are not is exercising discretion, no matter how precise the rule seems to be that they are applying, and no matter how rigorous the checks are on their decision. Discretion cannot be eliminated, but neither can it be easily structured or confined, at least not by rules.

So whether we like it or not, discretion is here to stay - what is important is how it is used. The focus here is how those enforcing regulation use the discretion they inevitably have in exercising their enforcement functions. The activities of 'street level bureaucrats' are important for the enforcement process for they act as its gatekeepers. As Diver has argued, '[t]heir decisions concerning what activities to inspect, what evidence to examine, what inferences to draw, and what action to take determine the composition of cases available for enforcement proceedings and constrain the practical range of prosecutorial policies' (Diver 1980; Barret and Fudge 1981; Rowan-Robinson et al 1990). But their activities have a wider significance. Understanding how enforcement officials exercise their functions is central to an understanding of how any regulatory system operates, for as many have noted regulation is not a product of legislators or those who write regulatory rules, rather it is the product of interactions between regulators, regulatees, and the wider community interested in the regulatory project (Olsen 1992; J. Black 1998; Meidinger 1987; Hawkins 1984).

The next section therefore considers how enforcement officers exercise their functions: how they construct definitions of compliance, and how they decide what to do in cases of non-compliance. The third section discusses how enforcement officers should be acting: what role should there be for the exercise of sanctions, what role for non-sanctioning responses to non-compliance. The final section considers whether and how the actions of enforcement officials could be influenced so as to bring their actual practice closer to the normative ideals. Whilst some general proposals can be made, it is unlikely that one solution will be appropriate for the multitude of regulators that exist, or that the solution will be simple. For three messages ring out loud and clear from the plethora of suggestions on regulatory reform both within and outside government: debates about reform should be grounded in as clear an understanding as possible of the nature of what it is that is seen to be in need of reform; they should be responsive to that particular context, and that to complex and difficult problems there are always simple solutions, and they are always wrong.

Discretion on the ground: how enforcement officials make decisions

Enforcement approaches and determining compliance

Enforcement is not just about gaining compliance with the law, it is about determining what compliance is. Many of the ethnographic studies suggest that what counts as 'compliance' is not simply a matter of mere conformity to a rule. Rather compliance is a matter of interpretation: interpretation of rules, and interpretation of facts (Hawkins 1984; Hutter 1997; J. Black, 1997). Indeed the rule may not be central to the definition of compliance that officials construct. Officials may have a view of full compliance, of a state which they are aiming towards, which may be above

the legal minimum, and certainly is in accordance with the purpose of the rule rather than its letter: substantive as opposed to rule compliance (Parker 1999b). They may therefore treat conduct which is conformity with the rule as unacceptable compliance, either because there is not compliance with the spirit of the law, or where although the minimum legal conditions have been met, they feel more could be done (Parker 1999b; Hutter 1997: 80).

Further, it is not always the case that non-conformity to a rule will be treated as non-compliance with respect to which action needs to be taken (though as discussed below the willingness of agencies to use sanctions in instances of non-compliance varies). Compliance has been characterised as a process rather than an event; a continuing effort towards the attainment of a goal as much as attaining the goal itself (Hawkins 1984: 110, 126; Hutter 1997: 13). That process is open-ended, long term, evolving from the interactions of several groups and factors, occurring over time, involving endless negotiation and characterised by a gradual progression of increasing standards (Di Mento 1986, Hutter 1997, Manning 1988; Hawkins 1984: ch 6; Edelman et al 1991). Most often, then, officials work with temporary definitions of compliance: states of affairs or conduct which are less than full compliance but which are tolerable for a certain period of time (Hutter, 1997: 80-85, Hawkins 1984: 108-109; 125; Richardson et al 1983: 152-4). Alternatively, or in addition, enforcement officers may accept conduct, practices or states of affairs that whilst not in conformity with the rules attain the overall purposes of the regulatory system (what Hutter calls 'sanctioned non-compliance' (Hutter 1997: 81-83). They may also adopt an aggregate definition of compliance, overlooking occasional infractions if on the whole the regulatee has complied (Hutter 1997: 83; Hawkins 1984: 27).

There is a considerable body of empirical research on how enforcement officials seek to gain compliance, and in particular what action they will take with respect to non-compliance (eg Carson 1970, 1982; Cranston 1979; Kagan 1978; Bardach and Kagan 1982; Richardson et al 1983; Hawkins 1984; Hawkins and Thomas (eds) 1984; Vogel 1984; Gunningham 1987; Rees 1988; Hutter 1988; 1997, Rowan-Robinson et al 1990; Braithwaite 1985; Grabovsky and Braithwaite 1986; Baldwin 1990, 1995; Yeager 1991; Gray and Scholz 1991). One of the early concerns was to compare how the activities of those enforcing laws regulating the operations of businesses differed from or were similar to the activities of the police (eg Kagan 1989). Two 'ideal types' of enforcement action were developed: a 'sanctioning' or 'deterrence' approach based on detection, apprehension and punishment, and a 'compliance' or 'accommodative' approach based on bargaining, conciliation and negotiation (Reiss 1984; Hawkins 1984). The sanctioning or deterrence approach is characterised by a central concern with securing punishment for breaching the rules. Future compliance with the rule may be a by-product of enforcement action, but it is not the central motivation. The motivation is to secure punishment, either for retribution and / or for a broader utilitarian purpose (Hawkins 1984; Hutter 1997). The style is accusatory and adversarial. The questions are whether a law has been broken and whether an offender can be detected. Enforcement is retrospective - determining the harm done, detecting the law breaker and fixing a sanction (Hawkins 1984: 4-5). In contrast, the compliance approach is characterised by a central concern with attaining the goals of the regulatory system. Securing repair of harm done and future compliance are the core motivations. The aim is to prevent harm rather than punish an evil. The style is conciliatory and relies on persuasion, education, negotiation and bargaining to attain results. Enforcement is prospective - responding to a problem and negotiating future conformity to standards that are administratively determined. (ibid).

Whilst a useful heuristic device, the division of enforcement styles into just two could never hope to capture the full range of enforcement approaches that exist. The distinction between 'compliance' and 'sanctioning' approaches focuses on only one variable: the use of punishment in enforcement, and asks simply is punishment used routinely for breaches of rules, or not. It therefore fails to allow

for any distinctions in approaches which do not use punishments routinely, but which may still vary in their deployment of them. Nor does it focus on other variables which may be part of an enforcement approach, for example the extent to which enforcement officers adopt 'legalistic' or narrow interpretations of rules (Bardach and Kagan, 1982), whether they are reactive or proactive, the number and frequency of inspections (see Gray and Scholz 1991), the manner in which inspections are conducted, or other differences in control strategies used by enforcement officers: education and advice, conciliation and mediation, and compensation and redress (on which see further D. Black 1976).

There have therefore been various refinements which have identified gradations of approach between the two poles of compliance and deterrence, though these have again focused principally on the use of sanctions. Hutter, for example, has further sub-divided the 'compliance' approach into 'persuasive' and 'insistent' (Hutter 1988: 131-176, Hutter 1997: 16) based on the readiness to use sanctions. In the persuasive approach, officials use measures falling far short of sanctions to secure compliance: education, coaxing, cajoling, persuading, patiently explaining rationales for law and means for compliance, and regarding the process of securing compliance to be an open-ended and long-term venture. In a more insistent approach, education and negotiation play a similar role, but officials are less benevolent and less flexible. There are fairly defined limits to their tolerance and they are more likely to deploy sanctions if compliance is not forthcoming, though the main aim of those sanctions is again to secure compliance rather than to exact retribution (Bardach and Kagan's ideal of 'flexible enforcement' (Bardach and Kagan 1982)).

Probably the most extensive attempt to develop a typology of enforcement approaches across a range of enforcement agencies is that of Braithwaite, Grabovsky and Walker (1987). In their study of 101 Australian regulatory agencies, they developed a typology of agencies on the basis of 127 variables grouped into seven categories: those relating to the structure of the agency, its powers, regulatory policy, enforcement practices, attitudes of enforcement officials, the structure of the regulated industry, and miscellaneous variables, eg the age of the agency. Based on an analysis of a sub-group of 39 variables (those relating to agency policy and practice), they identified seven dominant enforcement 'types': conciliators, benign big guns, diagnostic inspectorates, detached token enforcers, detached modest enforcers, token enforcers and modest enforcers. *Conciliators* are agencies which reject any kind of enforcement model, relying instead on bringing parties together to resolve disputes as the way to achieve regulatory goals, and they are reactive: all the anti-discrimination agencies fell into this category. *Benign big guns* also adopt a reactive approach, and are agencies that have extremely severe sanctions at their disposal (eg licence revocation, product approval, powers to close down businesses), but very rarely use them. *Diagnostic inspectorates* were distinguished by their policies concerning the nature of inspections. Inspectorates were decentralised, the approach was proactive, inspectors had a high degree of discretion, they were concerned to maintain co-operative relationships, offered advice and education on how to achieve compliance, and encouraged industry self regulation. *Token enforcers* in contrast were also proactive, but they were not concerned to educate, nor particularly to sanction. Instead inspections were perfunctory rulebook inspections: 'regulation by going through the motions' (Braithwaite et al, 1987: 340). *Detached token enforcers* were similar, but less inclined to build up any relationship with the firm, as were detached modest enforcers. Finally, *modest enforcers* were those that adopted a sanctioning approach: they were much more likely than any of the others to be punitive, prosecuting or imposing administrative sanctions.

Braithwaite et al concluded from their study that most Australian enforcement agencies were 'token enforcers', they performed perfunctory, rulebook inspections, and imposed occasional, perfunctory sanctions. They were neither aggressively adversarial nor captured: they mattered too little to be worth capturing (Braithwaite et al, 1987). In contrast, ethnographic studies of enforcement officials

in the UK (which have been conducted principally of officials enforcing environmental and health and safety law) have found that most would be in Braithwaite et al's terms 'diagnostic inspectorates': adopting proactive inspection policies, orientated to gaining compliance through negotiation, persuasion, education, advice and bargaining, though as Hutter noted some may use sanctions more than others (Hutter, 1988; 1987; Hawkins 1984; Richardson et al 1983; Rowan-Robinson 1990; Baldwin 1990; 1994; Genn, 1993; see also Gunningham 1987 (Australia)). In contrast, studies in the US show that the approach adopted there tends to be both more punitive and more legalistic: 'modest enforcers; or regulation 'going by the book' (Bardach and Kagan 1982; Vogel 1984; Kelman 1981; 1984), although variations can occur between agencies and branches of the same agency (Shover et al 1984).

Those generalisations are interesting, but they are by their nature broad brush. They do not reveal or explain the variations of practice which exist within any one categorisation: when diagnostic inspectors, for example, seek to educate, when to bargain, when to impose a sanction, nor do they get us very far in understanding why different categorisations arise: what factors lead one agency to be a token enforcer and another a diagnostic inspector? Predictive models have in fact so far proved elusive; instead there is a rich body of evidence gathered on the different factors that influence the enforcement approach that officials in fact adopt. These studies are largely individual studies of agency enforcement, and of agencies which have a large body of enforcement officers which go out into firms to inspect business processes: health and safety regulation, and environmental regulation. Their activities may be very different from agencies enforcing competition law or compliance with utilities regulation, for example, where work is likely to be more team based and individual case officers may have far less operational autonomy than the individual inspector when he visits business premises. We should be wary then of extrapolating too much from the empirical work to produce generalisations, but that work nevertheless gives a good insight to what enforcement consists of at least in particular types of regulatory situations.

Factors influencing the enforcement approach adopted

Most are in agreement, therefore, that enforcement approaches are far more variable than the dichotomy of sanctioning and compliance approaches suggests, and that compliance is on the whole both a construct and an ongoing process. How enforcement officers construct working definitions of compliance and the approach they adopt to non-compliance is in turn the product of the complex interaction of a range of factors. The impact that each of those factors has, its weight, and how it interacts with other factors, are matters which are likely to vary in space and time. As noted, we should be wary of taking the study of one agency, or even several studies of agencies involved in enforcing the same type of regulation, and generalising those findings across all agencies. Moreover, as noted, those studies have been principally orientated to determining when sanctions will be imposed rather than to identifying and explaining other variations in enforcement approach. Nonetheless, whilst we might not be able to say everything there is to say, we need not remain mute. Sufficient studies have been done to allow at least identification of those factors which may be most influential in affecting when sanctions will be used, though as noted their exact impact is likely to vary between agencies and regulatory contexts. Those factors are the legal framework of the regulation including the sanctions at the agency's disposal, the nature of the interactions between regulator and regulatee, the type of breach, the type of firm, internal bureaucratic leadership, organisation and culture, the personal backgrounds, experiences and attitudes of enforcement officers themselves, and the broad political, economic, social and moral context of the regulation.

Legal framework

The substance and form of rules

The clarity of the regulatory goals, the inherent nature of the regulatory obligations being imposed, and the form of the rules used can all have an influence on the type of regulatory approach adopted. Some have suggested that the more that the overall goals of the regulatory system are clear, few in number, and non-contradictory, the more regulators may be prepared to impose sanctions for their non compliance (Hawkins and Thomas, 1984b: 10). Further, the inherent nature of the activities being regulated and the regulatory obligations being imposed may necessitate an approach which is orientated more to negotiation than to punishment. The behaviour that is being regulated is continuous and on-going, and the requirements that are imposed require a positive accomplishment, eg the installation of special technology, operational systems, training, and need time and money for their attainment. They are not the types of goals, in other words, with which compliance can be instant. In such a regulatory context, enforcement officers must display patience and tolerance rather than legal authority for the goal is not to punish but to secure long term change. (Hawkins 1984: 197).

Not just the substance but the types of rules that are used to define the regulatory requirements may affect which approach is used (Hawkins 1989; Baldwin 1990, 1995), although rule type seems more to help or hinder the adoption of a particular approach rather than be a strong causal factor as to which is adopted. The use of general rules which are vague as to manner, place, or time, and / or which impose evaluative standards ('reasonable', 'best practicable', 'suitable') implicitly confer discretion to those applying the rule (not just enforcement officials) to make a judgement as to its application in any individual circumstance. They make it easier for a compliance approach to be adopted, lending themselves to persuasion and negotiation as to exactly what conduct is required, although general rules may in fact develop a quite particular interpretation, and can in fact facilitate sanctioning in instances where detailed rules do not cover the conduct in question (J. Black, 1999). In contrast, the use of precise, 'bright line' rules can facilitate a sanctioning approach if it is clear on the face of the rule what conduct is required or prohibited. However, a mass of detailed rules can in fact increase discretion rather than reduce it (J. Black, 1995, 1997), and rather than facilitate enforcement it can bring it to a halt as there are simply too many rules to monitor (Rowan-Robinson, 1990).

Sanctions available

The sanctions available and who has the decision whether or not to impose them may also influence the type of approach adopted. Many studies suggest that the nature of sanctions available, and whether it is the agency or another body that imposes them, can have an impact on the decision by officials of whether to escalate the case up through the agency with the aim of seeking a sanction, although again the range of sanctions available seems to help or hinder particular enforcement approaches rather than be a strong driving factor in determining the approach adopted. If enforcement officials perceive the imposition of sanctions to be a sign of failure, they will be uninterested in having greater sanctioning powers (Gunningham 1987).

That said, if there are only weak sanctions available (eg prosecution with a low conviction rate and / or low level of fine), then enforcement officers are unlikely to seek their imposition (Richardson et al 1983; Hawkins 1984; Hutter 1988, 1997; Rowan-Robinson et al 1990), although that does not mean that they will not use the threat of the sanction to gain compliance. Indeed, the threat of sanctions, particularly of prosecution, plays a significant role in the ability to adopt a persuasive approach (Hawkins 1984; Grabovsky and Braithwaite 1986; Rowan-Robinson et al 1990, and further below). But threats only work if there is fear that the consequence threatened will actually be as bad as feared. Thus where the sanctions threatened are in fact weak, then officers are unlikely

to seek to use them for their impact will be revealed to be derisory and the threat of emptied of any content. Bluff only works if it is never called.

However, having what on the face of it appear to be 'strong' sanctions, prosecution with a high rate of conviction and / or a high fine or licence revocation, for example, does not mean that sanctions will be frequently sought or imposed. Officials with strong powers can still adopt compliance approaches even where there is no evidence of capture, and indeed may be more enabled to do so as all participants know that they bargain in the shadow of a large stick (Grabovsky and Braithwaite 1986). The threat of sanction still has to be real, however. Just as sanctions that are too weak pose no threat, neither do those that are too strong. The strength of a threat may depend not just on the type of sanction but of the context of its use: the threat of individual licence revocation may be entirely credible in the context of those licenced to issue consumer credit for example, where there are many providers and the wider social and economic consequences of revocation are small, but the threat of licence revocation for British Telecom has far less credibility, for there is only the very remotest possibility that it would be used. 'Strong' sanctions may thus lack any credibility. Further, the sanction may simply be too strong to merit being used for the preponderance of offences - it is simply disproportionate. Finally, as discussed further below, the greater the sanction, particularly if it is a criminal one, the greater the moral culpability that enforcers may see to be a precondition for its use, and so the less it may be used.

Where a range of sanctions is available, studies suggest that the choice is likely to be determined by the facility with which it can be used and by its effectiveness in securing compliance, meting out punishment and / or deterring offenders (eg Rowan-Robinson 1990). Their study of enforcement practices of several UK agencies suggests that where there are administrative sanctions, the level of prosecutions is reduced, and that one of the key influences on whether to prosecute is whether other sanctions are available (Rowan-Robinson 1990: 244). Where they are available, administrative sanctions are far more likely to be used than criminal prosecution.

Administrative sanctions may be preferred to prosecutions for reasons additional to those of effectiveness and expeditiousness, namely those of trust and interpretive control. The decision to escalate the case up the organisation and to seek prosecution can be a complicated one for the enforcement officer. Moving the case up the agency means the officer loses control of it, just as prosecution means the agency loses control. Studies by Hutter and Hawkins indicate that enforcement officers' perception that magistrates would be overly sympathetic to the regulated (eg because magistrates were themselves farmers or businessmen or moved in the same social circles as the regulated) was a contributory factor in the decision not to seek prosecution (Hutter 1988: 72-74; Hawkins 1984). The concern is not just that the agency will lose. It is that the agency risks an unhelpful precedent being set, even if it wins on the facts. That is, the agency runs the risk that the interpretation of the rules that the court will adopt will authorise a lower standard of behaviour than the agency would wish (see further Hawkins 1984, Hutter 1997, J. Black 1997a, 1999; McBarnet and Whelan 1999). It may be, therefore, that enforcement officers would be more ready to use sanctions if they were to be imposed by the regulator (even a separate division) than by a court. Or more precisely, that they would be imposed by a body that the enforcers felt it could trust, and which was part of the same interpretive community (see J. Black, 1997a, 1999). However, even if the decision to impose a sanction is a regulatory or administrative one, the potential for that decision to be appealed to a court or tribunal or be subject to review may have the same effect: the agency loses control of the interpretive framework, and so may be disinclined to pursue formal action (see further below). Of course it may be forced into such action by a hostile regulatee who consistently refuses to play the negotiation game.

The interactions between enforcer and regulatee

The design of the legal system is relevant to the type of approach adopted, but it is far from determinative. Other factors have a far more direct impact. As studies of enforcement and of rule making have emphasised, regulation in practice is the product of the day to day interactions of regulators and regulatees (Hawkins 1984; Hutter 1988, 1997, Baldwin 1990, 1995; J. Black 1998; Parker 1999a). The perceptions that enforcement officers have of firms and the nature of their relationship with them are key factors in determining the enforcement approach adopted.

The characterisation of the regulatee

In contrast to the rather indeterminate role of the general regulatory framework, how enforcement officers characterise the regulatee appears to have a direct and significant effect on the enforcement approach adopted. Hawkins, for example, argues that it is the image of the polluter that can be central to the construction of notion of compliance and for constructing notions of responsibility for non-compliance, and thus for determining what the enforcement response should be (Hawkins 1984: 110-118). Developing typologies of regulatory firms is a common practice for enforcement officials, and for the academics who study them (Bardach and Kagan 1982; Hawkins 1984; Kagan and Scholz 1984; Genn 1993; Baldwin 1990, 1995; Hutter 1997; Haines 1998). Best known is Kagan and Scholz's tripartite classification of firms into amoral calculators, political citizens, and the organisationally incompetent (Kagan and Scholz 1984). *Amoral calculators* are motivated entirely by profit-seeking, and their decision of whether or not to comply is an economic one based on the calculation of the relative costs of non-compliance (sanction and probability of detection) and the profits to be gained through non-compliance. *Political citizens* are firms that are ordinarily inclined to comply with the law, either because they agree with its goals or just because it is the law, but that willingness is contingent. At least some law breaking stems from principled disagreement with the regulation or with requirements they consider to be arbitrary or unreasonable (corresponding to Baldwin's well intentioned and well informed: 1995: 151). The *organisationally incompetent* are also inclined to comply, but fail to do so due to weaknesses in internal controls and / or inadequate knowledge of the law (Baldwin's well intentioned but ill informed: 1995: 149). Worth adding is probably a fourth category: the *irrational non-compliers*, who are malicious, or ill intentioned and ill informed (eg Hawkins 1984: 110-118; Baldwin 1995: 151): those who deliberately do not comply, but not because of a rational economic calculation but in symbolic rejection of the agency's authority.

Sociological and socio-legal research on enforcement has shown that regulatees are not necessarily 'amoral calculators' (Kagan and Scholz 1984) or indeed irrational non-compliers, rather they are on the whole either political citizens or organisationally incompetent. Ayres and Braithwaite have cautioned, however, against dividing firms into particular types. Firms, they argue, may be both amoral calculators and corporate citizens, either across space with different divisions, area offices or management levels having different motivations, or over time: the firm that may be a responsible citizen today may next month be an amoral calculator. Indeed, the same people may at different times, or indeed the same time, be motivated by profit maximisation and by a sense of social responsibility (Ayres and Braithwaite 1992, 30-35).

Enforcement officers are perhaps more sensitive to this variability than academics, and studies show that they discriminate between individuals, between sites, as well as between firms (eg Hutter 1997). Characterisations are in practice developed of particular firms, of types of firm, and of individuals within firms. Those characterisations provide the frame in which assessments are made of whether or not particular instances of non compliance are due to negligence, wilfulness, or are simply accidents, and they affect the enforcement strategy which is used in response (Hawkins

1984; Hutter 1988, 1997; Kelman 1984; Kagan and Scholz 1984). Hawkins, for example, found that enforcement behaviour in pollution control was determined by the interplay of two factors: the nature of the deviance and a judgement of its wilfulness and avoidability (Hawkins 1984:105; Hawkins and Hutter 1993)

Given the importance of the enforcement officers' perceptions of the characteristics of the regulatee for the type of enforcement approach they are likely to adopt, it is important to consider what factors give rise to particular characterisations. Most significantly, probably, is the assessment of the regulatee's motivation to comply, which is in turn based on assessments of the company's commitment to the purposes of the regulation (eg Hawkins and Hutter 1993). These assessments are subjective, and are based on, for example, the time, money and energy devoted to achieving the goals of the regulation (eg health and safety, pollution control); the quality and attitude of staff towards compliance, both managers and employees; the company's ability to comply, judged both on the basis of its financial position and the degree of technical knowledge it possesses, the treatment of its workforce, staff turnover and the nature of any incentive structures in place. Many of these assessments are made on the basis of the firm's 'enforcement career' (Hawkins 1984). That is, the firm's past record of compliance, of honesty in its dealings with the agency, and of previous responsiveness to enforcement officers' demands (Hawkins 1984; Hutter, 1988, 1997, Richardson et al, 1983).

Characterisations are also constructed of particular types of firms and particular types of occupation. Occupations or businesses that are often characterised as being malicious or uncooperative are construction sites, clothing factories (Baldwin 1995: 151), scaffolders, contractors, second hand salesmen (Cranston 1979: 34; Hutter 1997: 180; Rowan-Robinson 1990: 232). Many of these are short term, itinerant or ephemeral businesses: in Hawkins' study, for example, enforcement officers expected small companies in urban areas, especially if they involved 'self made men', to be 'careless' or malicious and uncooperative. However, farmers were also expected to be particularly recalcitrant, though they were long standing businesses. (Hawkins 1984: 114). In contrast, those in professional jobs, (engineers, chemists, craftsmen) are often characterised as likely to comply (Hutter 1997: 180).

Characterisations can be developed through enforcement officer's own personal interactions with the firm over a period of time, or may be lodged in the institutional memory of the enforcement agency and communicated to those who have had no prior dealings with the firm (Hawkins and Hutter 1993). Characterisations can also shift over time: a firm can move from a positive to a negative assessment based on its past record of compliance. Characterisations of particular individuals encountered within those businesses, however, can override the characterisation given to the general group to which the firm belongs, though it will not necessarily negate a past history of non-compliance (ibid).

As noted, the typical characterisation adopted is that most firms are generally inclined to comply, and that the reasons for non compliance are ignorance and incompetence rather than deliberate intent (Cranston 1979, Hawkins 1984; Hutter 1988, 1997; Grabovsky and Braithwaite 1986; Gunningham 1987). Amoral calculators or irrational non-compliers tend to be considered to be a small minority of those regulated. However, characterisations are largely subjective assessments and how they are constructed may depend not so much on objective characteristics of the firm as on the way that either that firm or firms like it have been labelled in the past by the enforcement official or the agency, and on the officers' own background and experience and interpretation of their own role. Rather than the characterisation determine the enforcement response, the desired enforcement response may determine the characterisation (or at the least there is a reflexive relationship between the two). Richardson et al, for example, found a 'striking coincidence between

officers' own preference for a co-operative approach and their attribution of non-compliance to those causes that could best be dealt with by that approach' (Richardson et al, 1983: 188, see also Gunningham 1987). So characterisations, whilst important, interact with, and may be a product of, other factors such as relational distance, organisational ethos and personal backgrounds and preferences.

Relational distance

The largely subjective assessments on which characterisations are constructed are in turn linked to the nature of the relationship that exists between the enforcement official and the regulated firm, or the individuals within it with whom the official deals. A central theme in enforcement literature is the impact of this relationship on the enforcement approach adopted. Influential in this respect has been Black's notion of relational distance (D. Black 1976). He predicted that the quantity of law will vary directly with relational distance: the higher the distance, the more law will be used in the case of disputes, and vice versa. Relational distance can be measured in a range of ways, which include the scope, frequency and duration of interactions between people, the age of their relationship, and the nature and number of links between them in a social network (ibid: 4, 41).

Some have argued that at a generic level there is likely to be a relatively low relational distance between regulatory inspectors and regulated firms compared to that between policemen and offenders, at least those who are not dealing with serial offenders (eg prostitutes) due to the different nature of the deviance. If the law breaking is categorical, unproblematic, momentary, discrete, bounded and unpredictable in its location and distribution, then there is little or no relationship between the offender and the enforcer, which will tend to prompt a sanctioning strategy (Reiss 1984; Hawkins 1984). If, as in the case of most regulatory offences, the conduct is continuing, repetitive or episodic, and offences arise from states of affairs as much as specific acts, then the nature of the deviance provides for the development of social relationships between enforcer and potential deviant. Enforcement is then a serial, incremental and ongoing process, directed towards compliance (Hawkins 1984: 6).

More interesting for these purposes is not the differences between policing and regulatory enforcement but the variations in approach within the regulatory context. In that context, Black's work has been applied to the work of Australian regulatory agencies by Grabovsky and Braithwaite (1986). They hypothesised that

- (i) an agency with a high percentage of staff drawn from the regulated industries would prosecute less than those whose staff were recruited from elsewhere
- (ii) agencies which regulated a relatively few number of firms would prosecute less than those that regulated a high number
- (iii) agencies which regulated a homogenous industry sector would prosecute less than those that regulated a heterogenous sector
- (iv) agencies whose inspectors had frequent contact with the same firms would use less formal sanctions than those characterised by less personal contact.

Their study found support for all but the first of these hypotheses: the greater the relational distance, as measured on these bases, the greater the use of formal sanctions.

Support for Black's thesis has been found by other studies of social regulation, although the findings have not been identical to those of Grabovsky and Braithwaite. Contrary to their findings, shared professional experience (Hood et al 2000: 60-65) and involvement in the same social community (Hutter 1997) have been found to be relevant factors in reducing relational distance.

Further, Hood et al's study of regulation inside government failed to find a clear relationship between the formality of the regulatory approach and the number of regulatees (Hood et al 2000), though other studies of regulation of business have found that the lower the ratio of officers to regulatees, the less formal the relationship (see eg Hawkins and Hutter 1993). Nonetheless, most studies have found support for the thesis that the greater the frequency of contact and the higher the homogeneity of the regulated, the more conciliatory the regulatory approach.

The size of firm

Variations in the assessments of firm occur not just between different industry sectors but between small and large firms, and the enforcement approach adopted can vary accordingly. Most studies suggest that a sanctioning approach is more likely to be adopted towards small firms than large firms, even in the context of a predominantly compliance based approach. (Grabovsky and Braithwaite 1986, Yeager 1991, Clarke 1990, Cook 1989, Gunningham 1991, Levi 1987, Shapiro 1984 but contrast Hutter 1988, 187-8). Just why this is the case is disputed in the literature. Some suggestions are that a lower relational distance is likely to develop between enforcement officers and large firms simply because it is more likely that officers will have more frequent contact with larger firms. They are likely to be long term as opposed to transitory operators, so enforcement officers will build up relationships through repeated interactions (Reiss 1984; D. Black 1976; Grabovsky and Braithwaite 1986). Further, because they are larger, inspections will usually take longer, and because their activities are more complex they will be more reliant on firms for information, which they believe is more likely to be forthcoming if a compliance approach is adopted (Hawkins and Thomas 1984b; Yeager 1991; Hutter 1997). Others suggest the reasons are more deep rooted and lie in the nature of capitalist society (Carson 1979, 1982; Yeager 1991).

Most large firms are also characterised as being 'political citizens'. Enforcement officers tend to consider large firms have a greater motivation to comply because, for example, they are believed to be concerned about their reputation (Grabovsky and Braithwaite 1986, Hawkins 1984), and they are considered to have a greater financial resources and organisational capacity to comply than smaller firms. This may in part explain the adoption of a compliance based approach. However lower use of sanctions or formal requirements to act could also be because larger companies have greater capacity to challenge any requirements made by the officer or sanctions imposed. These challenges could be to the officer directly, and / or the company could simply negotiate with more senior regulatory officials, exert influence in the political process, and appeal any sanction imposed (Yeager 1991). Large firms also tend to be better informed about the regulatory requirements and about the consequences of non-compliance, and so are less susceptible to bluff (Hawkins 1984; Shover et al 1984), something that is relevant if the sanctions available are weak or non-credible.

Compliance costs

Also significant can be the officers' assessment of the cost of compliance and the financial state of the company (Richardson et al 1983; Hawkins 1984; Hutter 1988; Kagan 1989; Rowan-Robinson 1990). Perceptions of the company's financial state are often built on quite subjective grounds (eg the size of the cars in the company car park: Hutter, 1997). If the officer thinks the company can afford to take steps to achieve substantive compliance and has not yet done so, he or she will sanction temporary compliance only for a short period. In the context of a broadly compliance based approach, however, in absence of some other factor (serious breach, immediate risk of harm, persistent failure or moral blameworthiness), the officer is unlikely to seek a sanction for non-compliance.

Nature of the breach

As noted, the moral culpability of the breach and the nature of the consequences arising from it can be important factors prompting the use of sanctions in an otherwise broadly compliance based approach. If the officers perceive that the breach gives rise to an immediate risk to health, safety or the environment, for example, or a direct harm has already resulted, then substantive compliance would be demanded (Hawkins, 1984; Hutter, 1997: 97-98, Richardson et al, 1983 157-161), regardless of whether it is considered that any blameworthiness was involved. However: whether a 'one-off' breach will be considered an accident or as a deliberate breach will depend largely on whether it is assessed to be an accident or to have been deliberate. That assessment of moral culpability will in turn depend on the overall characterisation which has been constructed of the firm (Hawkins 1984: 105). If the breach is deemed to be accidental, then it is unlikely that a sanctioning approach will be used unless, as noted, the breach was severe in nature or consequences (Hawkins 1984: 202). Prosecutions for persistent failures will be brought even in the context of a heavily compliance based approach, but they are seen as failures, and tend to be sought in order to punish not the breach itself so much as the symbolic assault on the legitimacy of the regulatory authority that the persistent non-compliance represents (Hawkins 1984: 205).

Organisational factors

The nature of the interactions between enforcement officials and regulatees is thus critical to shaping the enforcement approach adopted by officers in particular circumstances. But those relationships are formed in a wider context which includes the organisational structures, resources, practices and culture of the regulatory body as well as the wider political, social and moral context.

Lack of resources is sometimes cited as the reason for a low level of prosecutions, but lack of prosecution is arguably more likely to arise because of the other factors indicated: characterisation of the firm, relational distance, nature of the firm, nature of the breach, and the social, political and moral context. As Rowan-Robinson et al have argued, agencies will allocate their resources in a way which they consider will best achieve their ends (on the whole). If they do not see themselves as prosecutors or prosecution is seen as a sign of failure, they will not allocate resources to prosecuting. If they consider that advantages could be gained from prosecution, for example by making an example of a firm, then they will prosecute (Rowan-Robinson 1990). That said, lack of resources may mean that officers spend less time educating or advising firms, and may adopt a more insistent approach simply because they have a heavy workload and it is quicker to issue a notice than spend hours explaining just what needs to be done (Hutter 1997), but lack of resources does not of itself explain a low level of prosecutions.

Internal policy documents as to what approaches should be followed are on their own unlikely to have much of an impact (Hutter 1988; Rowan-Robinson et al 1990). Nonetheless, certain organisational practices can affect the enforcement approach adopted. If the enforcement practice is to be reactive, to investigate only or predominantly on receipt of a complaint, or after an accident for example, then the agency may operate in a more formal manner than those agencies that are proactive in enforcement. Some studies have found that where agencies are investigating in response to a complaint then they are more likely to 'go by the book' (Hutter 1988; Cranston 1979; Kagan 1989; Kelman 1984), but others have found that agencies deliberately adopt a mediating or conciliatory stance in such cases (Silbey 1984), acting as conciliators rather than as modest enforcers.

Organisational practices can also affect the enforcement approach by increasing the relational distance between the firm and the agency. Frequent rotation of enforcement personnel to different

geographical areas or different firms prevents officers from building up long term relationships with particular regulated firms (for examples see eg Hutter 1997; Hood et al 2000). Relational distance can also be increased by including others in the enforcement process. This may be achieved by having internal rules that require decisions even on low level sanctions to be referred up to those who could override the field officer, by having all reports by field officers scrutinised by others, if or requiring other specialists to become involved in enforcement decisions (eg technical experts, lawyers, economists) (Rowan-Robinson et al 1990: 221-2). If the enforcement process is principally reactive, based on complaints received, then it could be by having to seek approval from another person or team before the case can be closed. These measures have the effect of reducing the autonomy of the individual enforcement officer, and perhaps of increasing the relational distance between the officer and the firm (although this may make little difference if the regulator as a body already has a close relational distance with the regulatee). To the extent that relational distance is increased, however, a more formal and sanctioning approach may be adopted, though on the other hand it may be that more senior officials are more easily persuaded by firms that sanctions should not be imposed than more 'hardened' enforcement officers (Hawkins 1984). Further, it may not be the case that relational distance is lower between field officers and firms and increases as one moves up the agency hierarchy: it may be that those more senior in the agency enjoy closer personal relational distance with senior management in the firm than do individual enforcement officers on the ground.

Internal performance measures linked to pay and promotion can have an impact on the way that officers operate, although it may be more accurate to say that it affects how they present their work to their superiors than the job they actually do. Officers are likely to focus on those aspects of their work that will receive approval from their managers. But if the measurement of good performance is linked to the number of formal actions taken (samples taken, notices issued, warnings given etc), then officers are likely to behave accordingly (see eg Hawkins 1984; Hutter 1997).

But just as significant as organisational practices or policies is organisational culture, which may or may not be manifested or reinforced in policy statements or organisational practices. All the studies conducted show that socialisation into the norms of the organisation, peer pressure and images of what constituted a 'good' enforcement officer play an extremely strong role in affecting the type of enforcement approach an officer is likely to have (eg Hutter 1988, 1997; Richardson et al, 1983; Hawkins 1984; Kelman 1984; Shover et al 1984; Hood et al, 2000).

Internal leadership within the organisation can play a strong role in defining that organisational culture (Hawkins and Thomas 1984b; Kagan 1989). But again it is important to recognise internal variations and tensions within the enforcement agency itself. There may be tensions within the organisation as to what a 'good' enforcement officer should be doing. These differences may exist horizontally, between regional offices of the same agency (eg Shover et al 1984), or vertically, between different layers of management. Whilst field level officers are concerned to maintain good long term relations through co-operation and bargaining and may see the most important part of their job to be out in the field, senior management may be concerned with ensuring that the agency is taking some visible action and require proof of effectiveness, and middle management may be concerned to ensure that officers are being active and require evidence through written records, monitoring time spent on inspections and time in the office (activity is easier to measure than effectiveness: Hawkins and Thomas 1984b: 19). As Hutter and Manning conclude: 'The pattern of enforcement is, at least in part, a result of the dialectic between management, with their concern for overall performance, compliance and the organisational mandate, middle management whose aim is to control the inspectors and their practices, and the inspectors who are inclined to see their task as exercising an immediate face-to-face responsibility and resolving culpability on their particular "patch" or "turf"' (Hutter and Manning 1990: 127).

Personal backgrounds

Personal inclinations are also important, and are related to officer's own experiences and expertise, their relationship with firms, and their own personal values (Richardson et al 1983, Hutter 1988, 1997). Their own moral norms or previous experiences of involvement in accidents, for example, is likely to lead to greater focus on what had caused that particular accident and a less flexible approach adopted with respect to it (Hutter 1997; Hutter and Lloyd Bostock 1990). Professional background may have an effect, but the evidence is mixed (see eg Richardson et al 1983; Meidinger 1987). What is important is the officer's own perception of his or her role and of the organisation's expectations (eg Rowan-Robinson 1990; Gunningham 1987; Hawkins 1984). In their study of pollution control, for example, Richardson et al found that the inspectors all had professional backgrounds in science, and saw themselves as able to offer technical advice, assistance and education, not as policemen. As Richardson et al summarise: 'The officers' behaviour was restricted by professional expectations, bureaucratic demands and personal moral norms. An action in disregard of any of these would be considered 'illegitimate' in some sense: a breach of professional expectations would be unprofessional; of bureaucratic demands, irregular; and of moral norms, immoral' (Richardson et al, 1983: 190).

The context of regulation

'Regulators take their cues... from the shifting demands and expectations of those whose continuing contributions of fiscal resources, political legitimacy and legal authority simultaneously define and facilitate their mission' (Diver 1980). Regulatory agencies operate in a context in which political demands, media attention, economic conditions, social concern and moral legitimacy exert complex and often contradictory pressures. Those pressures are manifested at all levels: from individual inspectors through to agency executives. They can affect enforcement priorities, inspection procedures, and its sanctioning activity (eg Shover et al 1984, Hutter and Manning 1990, Hutter 1997). Exactly how those pressures are rationalised within the agency, and how the inevitable tensions between them are or are not resolved will vary from agency to agency and from time to time. Needless to say, the more the agency is in the political or media spotlight, the more they are likely to adopt a sanctioning approach so as to indicate to those audiences that 'something is being done', though the approach might last only as long as the spotlight lingers.

Most of the time there is low public concern about the regulation, and considerable moral ambivalence as to the merits of the regulatory goals. The moral ambivalence surrounding regulation has been cited by Hawkins as the single most important factor which gives rise to a compliance approach (Hawkins 1984; see also Rowan-Robinson et al but contrast Kelman 1984: 136-7: social consensus gives rise to compliance approaches). Enforcement officials 'borrow' on the legitimacy of the law they are charged with implementing; absent that legitimacy, their position is precarious indeed. For the preponderance of enforcement activity, Hawkins argues that the compliance strategy will dominate, for it is a means of sustaining the consent of the regulated when there is ambivalence about an enforcement agency's legal mandate: 'bargaining is not only adjudged a more efficient means to attain the ends of regulation than formal enforcement of the rules, bargaining is, ultimately, morally compelled' (Hawkins 1984: 8-13, 127-128).

Summary

What determines how enforcement officers will use powers that they have been given is thus a complicated set of interactions between a range of factors: the legal framework, the characterisation of the regulatees, the relational distance between regulator and regulatee, the type of firm, the nature of the breach, organisational structures and norms, personal backgrounds and attitudes, and the broader political, social and moral context in which the regulation sits. Some indeed have argued that deeper structural factors relating to the development of trust systems and the growth and complexity of organisational life mean that compliance based approaches are likely to be the norm (Reiss 1984; Rock 1995). Before we move on to consider how, if at all, the legal system and bureaucratic organisation can hope to affect how those powers are used, we need to consider what the end goal should be: how should enforcement officers be using their powers?

How should enforcement officers enforce regulation?

Principles of enforcement

How enforcement officers should exercise their powers is at one level a question of legality and political morality, and at another one of effectiveness in achieving the goals of the regulatory system, however open ended, incoherent and ambiguous they may be. In terms of legality, then clearly officers have to act in accordance with their legal mandate and with those principles that apply to all those acting in the course of public office. In terms of political morality, then as Galligan has stated, exercises of discretion should comply with standards of rationality, purposiveness and morality (Galligan 1986: 5-6) which in turn may or may not be completely expressed in the duties imposed on them in public law. Drawing on those principles to guide enforcement action would suggest that enforcement officers should act, for example, legally, consistently, rationally, non-discriminately, proportionately, transparently, accessibly, expertly, and in accordance with the principles of due process.

Effectiveness of enforcement

In the debate on effectiveness, academic attention has again focused principally on the question of when to impose sanctions. Some work has been done which has attempted to assess the relative effectiveness of inspections, for example their frequency and duration (Gray and Scholz 1991), and to consider what enforcement activities should focus on (examination of records, practices, assessments of internal compliance systems: eg Parker 2000 forthcoming), but for the most part the debate has followed the lines of whether agencies should adopt a policy of punishment or persuasion (see Braithwaite 1985).

The normative debate thus follows the compliance / deterrence dichotomy (for heated discussion see Hawkins 1990; Pierce and Tombs 1990). Arguments usually put in favour of adopting a compliance approach are that it involves an efficient use of resources (persuasion is cheap), it will elicit a more co-operative approach from the regulatee, more information from the firm is likely to be forthcoming about its practices and possibly about areas of non-compliance, and it will engage the firm in decisions as to how best to act to secure compliance. Moreover, a compliance approach allows adjustment for the over-inclusiveness of rules (and thus can be used where the rules are not appropriate, for example because of poor design or a rapidly changing regulatory context). It allows identification of the best ways to improve performance, and because learning is involved; can improve compliance before harm results, so improving responses to risks. On the other hand, a

compliance approach may be exploited by regulatees, leading to ineffective regulation and indicating regulatory capture.

Arguments in favour of a deterrence approach are that it indicates a tough regulatory approach in which non-compliance is treated as unacceptable, so reinforcing and giving effect to social disapproval and increasing social pressures to comply. The combination of regulatory and social pressure then makes it rational for firms to give a higher priority to compliance and forcing a change in management culture and systems. On the other hand, an approach in which rules are given a narrow interpretation and in which sanctions are readily imposed for breach is to practice 'regulatory unreasonableness' (Bardach and Kagan 1982), which alienates regulatees, and does not capture the advantages of the compliance approach. Indeed the advantages claimed for the sanctioning approach may themselves be optimistic. Scholz, for example, has argued that the deterrence theory on which it is based is only valid when a number of assumptions are true: corporations are fully informed utility maximisers, legal statutes unambiguously define misbehaviour; legal punishment provides the primary incentive for corporate compliance; and enforcement agents optimally detect and punish behaviour given available resources. These assumptions are rarely if ever true. (Scholz 1997). Not only is the strategy likely to be of limited effect, it may have significant costs associated with it, for it can turn dispositions to comply into dispositions to resist. Punitive approaches will thus result in a game of regulatory 'cat and mouse', with firms engaging in creative compliance and regulators trying to devise more and more specific rules to block loopholes (Bardach and Kagan 1982; Ayres and Braithwaite 1992; Braithwaite 1985, Makkai and Braithwaite 1993, 1994a 1994b; Parker 1997a; McBarnet and Whelan 1991, 1999).

The more profitable turn in the debate has been to consider which combination of strategies is most appropriate in which circumstances. As Braithwaite has said, the issue is not whether to punish or persuade, but when to punish and when to persuade (Braithwaite 1985). The answer that most academics at least have come to is that the enforcement strategy should match the firm's motivation and capacity to comply. Most adopt variations on Kagan and Scholz's argument (Kagan and Scholz 1984; Baldwin 1990, 1995; Ayres and Braithwaite 1992). For *amoral calculators* (and we may add the *irrational non-compliers*), regulators should act as *policemen*, adopting a sanctioning approach. For *political citizens*, the regulator should act as a *politicians*. They should be ready to compromise on values, adjust regulations to changing circumstances, but also provide leadership in developing and persuading businesses to adopt socially acceptable solutions - a compliance based model, or regulatory reasonableness (Bardach and Kagan 1982). For the *organisationally incompetent*, regulators should act as *consultants*. On finding non-compliance they should educate and advise: analyse the causes, locate weaknesses in internal controls and point out cost-effective ways of complying. There are dangers, however, in matching the wrong strategy to the wrong firm. Approaches based on persuasion and education will be exploited by amoral calculators. Sanctioning approaches will alienate political citizens and the organisationally incompetent. Neither sanctioning nor persuasion approaches will educate the organisationally incompetent in how to improve their capacity to comply.

Both Kagan and Scholz and Braithwaite have therefore argued that rather than try to predict in advance what type of firm the enforcer is dealing with, the cue should be taken from the firm itself. Inspectors should be prepared to shift from strict policemen, to politician and to consultant and back again according to their analysis of a particular case (Kagan and Scholz 1987: 87, see also Baldwin 1995). Co-operation with the regulator can be reciprocated; deviance should be responded to with a move to a sanctioning approach, and a move back to co-operation when the firm indicates that it will now comply (see also Ayres and Braithwaite 1992).

Kagan and Scholz's propositions have in effect been further developed into proposals for regulatory reform both as to the sanctions that should be available (Ayres and Braithwaite 1992) and as to the type of rules that should be used for each strategy (Baldwin 1995). The latter will be addressed further below, but it is worth dwelling here on the proposals for the range and deployment of sanctions. Ayres and Braithwaite argue that the 'tit for tat' strategy of matching regulatory response to corporate behaviour should be combined with the deployment of a pyramid of sanctions, and a pyramid of regulatory strategies. The more the regulated firm refuses to comply, the greater the sanction that should be adopted and the more intrusive the regulatory strategy that should be introduced. The two pyramids are discussed as if they were one by Ayres and Braithwaite, but it is worth separating them here, as we are concerned with what actions enforcement officers and regulatory agencies should be taking. Their options may be confined to deployment of sanctions, for they may not have the power to move up the strategies pyramid. Thus whilst it may be credible for an enforcement officer or his regulatory body to threaten to impose a particular sanction, it will not be credible for them to threaten to introduce a different regulatory strategy when that decision is one for the political executive or legislature to take. Further, the discussion of what regulatory strategy to adopt is a far broader question than that of which sanction to impose in a particular instance, and as Gunningham and Grabovsky have pointed out, Ayres and Braithwaite's strategy pyramid could be considerably expanded (Gunningham and Grabovsky, 1998).

Focusing then just on the sanctions pyramid, the idea is that regulators move progressively up the pyramid, starting with a persuasive approach which employs no sanctions and gradually progressing up through the hierarchy until the most severe is reached. Just what sanctions the pyramid will contain should vary with the area being regulated, the importance is there should be a range with differing degrees of severity. For as noted above, one of the reasons why sanctions are not sought or imposed is because they are either too weak in all cases, or disproportionate for most offences, or too strong to be credible. So what enforcers need is a range of credible sanctions that enable them to match sanction to the form of non-compliance. Moreover, Ayres and Braithwaite argue that the overall height of the pyramid is critical: regulators 'will be more able to speak softly when they carry big sticks' (Ayres and Braithwaite, 1992: 19).

The example of a sanctions pyramid that Ayres and Braithwaite give is set out below (fig 1).

Fig 1: Ayres and Braithwaite's sanctions pyramid (1992: 35)

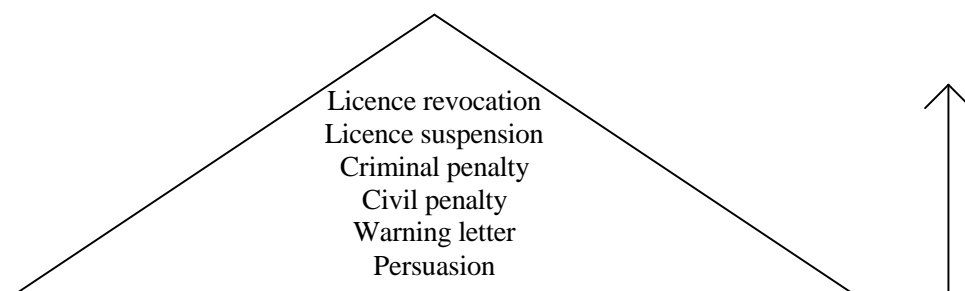
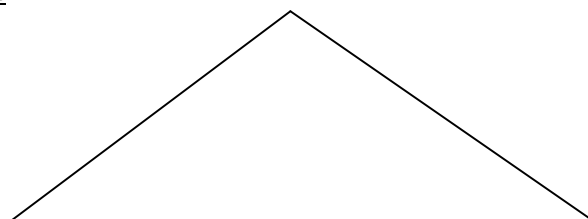


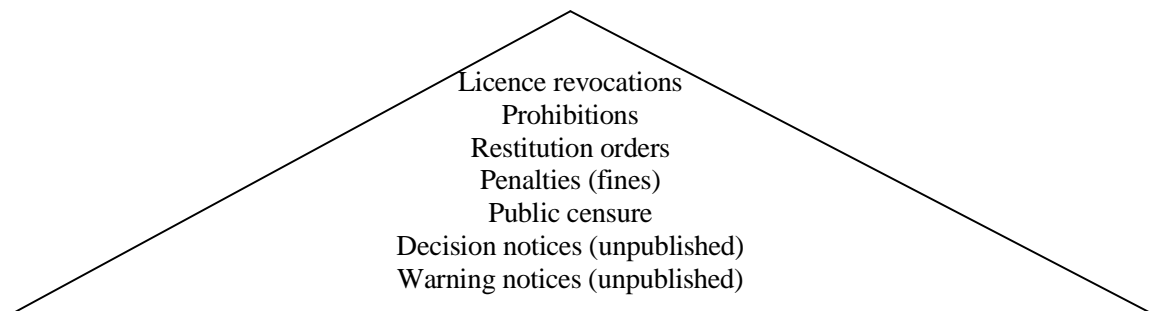
Fig 2: Example of a sanctions pyramid: sanctions available under Financial Services and Markets Act 2001*

Court imposed sanctions



Criminal offences
Civil liability and
private rights of action
Injunctions and restitution orders

Financial Services Authority imposed sanctions



*Note: Not all sanctions are available for all breaches of statutory or regulatory requirements, and sanctions are not normally mutually exclusive; arranging them in a hierarchy of severity is thus somewhat artificial (and indeed the suggested relative severity of each could be disputed).

In Ayres and Braithwaite's proposal, the pyramid is to be used in conjunction with the 'tit for tat' strategy: what the firm does, the regulator does. But the question still remains as to what the agency's starting move should be: should it always adopt a persuasive approach until the actions of the firm indicate it should move up, or should it always start higher up the pyramid, and be prepared to move down? Ayres and Braithwaite argue that because of the disadvantages of a punishment approach (expensive, counterproductive and unworkable in the long term), enforcers should always start off softly, initially adopting a co-operative approach, but should then respond to the firm's actions in kind. The greater the non-compliance, the greater the sanction imposed. However the enforcer should only use the minimal sanction necessary to secure compliance (the 'minimal sufficiency principle') (ibid., 40, 50), which is simply the principle of proportionality. This strategy, they argue, underlines the authority of the regulator, and supports the sense of fairness of responsible companies, who will comply but only if the dishonest are punished (ibid: 26, Bardach and Kagan 1984: 65-66). When the company does reform, then it can be 'rewarded' by the regulator reverting back to a co-operative approach, giving firms reasons to prefer their responsible selves to their economic or venal selves.

The proposal that regulators should have a range of sanctions of different degrees of severity and which are credible and appropriate for the regulatory context is to be supported. Nonetheless several criticisms can be made of the pyramid model and of the proposed strategy for their deployment. (It should be noted that those who never favour the use of compliance based approaches in any circumstances would automatically reject the bottom layer of the pyramid, persuasion, but that is not this author's view).

First, there is a tension between the 'tit for tat' strategy and the pyramid model. Whilst Ayres and Braithwaite argue that the regulator should respond in kind to the regulatee's actions ('tit for tat'), they also argue that the regulator should always move progressively up the pyramid. In other words, that the layers of the pyramid are lexically ordered: one can only move to the next level having completed the level below. But this seems unnecessarily restrictive. If the breach is severe, be it in its moral culpability and / or consequences, the tit for tat strategy suggests that the response should be equally severe. Less severe sanctions will be not be appropriate, and so there seems little point in escalating up through them. This seems to be the attitude of enforcement officers as well, for even in a predominantly compliance based enforcement approach they will in such cases target sanctions in accordance with the nature of the breach (eg Hawkins 1984; Richardson et al 1983; Baldwin 1995; Hutter 1997). So rather than seeing sanctions as lexically ordered in a pyramid up which one must escalate, it is perhaps better to see them as a range of tools from which enforcement officers can choose, subject always to the principles of proportionality and due process.

Second, the model is deceptively simple in that it assumes that regulators will be able to assess easily whether the firm is an amoral calculator, political citizen, organisationally incompetent or an irrational non-complier so as to adjust their enforcement approach accordingly (see also Haines 1997: 219). Tailoring the enforcement response to individual firms is also highly resource intensive; it demands skill, time and other resources that are likely to be in short supply. The strategy also requires certain structural conditions: at the least that the firm and the regulator are in a long term relationship that will enable the regulator to observe and assess the firm's actions over a period of time. But with transitory and itinerant operators, there is unlikely to be the opportunity to do so. Again, most empirical studies have found that enforcers usually find it appropriate with such firms to use a sanctioning approach straightaway, even in the context of an overall strategy based largely on persuasion (Hutter 1997, Hawkins 1984, Rowan-Robinson 1990; Gunningham and Grabovsky 1998). Further, it is unlikely, as Ayres and Braithwaite's own analysis suggests, that firms will be easily consignable to particular categories. Firms are complex organisations with multiple units and 'multiple selves'. Different parts of the same firm will at the same time have different motivations, as may the same parts at different times, or even at the same time.

For the 'tit for tat' strategy to work, moreover, not only does the regulator need to be able to interpret the firm's moves, the firm needs to be able to interpret the regulator's. So firms need to know that if they co-operate they will be 'rewarded' with a co-operative response, and that if they deviate, they will be sanctioned accordingly. This knowledge can only be gained through repeated interactions, or through knowing what other firms' experiences have been. The strategy is therefore unlikely to work for the irrational non-compliers, or the ill intentioned and ill informed, who by their nature have very little knowledge or awareness of the regulatory requirements, and are unlikely to know of the enforcement strategy. It is also unlikely to work with other firms unless there a sufficient degree of common understanding both as to the regulatory requirements and as to the enforcement game that the regulators are playing. As Parker has argued, regulatees and regulators must have enough in common to be able to understand when escalation up the pyramid will fit regulatory goals and when it is not necessary (Parker 1997a: 224).

Further, the proposal that regulators should always start off with a persuasive approach is based on the prediction that the costs of an inappropriate use of the persuasive approach are lower than the costs of the inappropriate use of the sanctioning approach: being taken for a ride by some is less costly than alienating some. This may or may not be true, whether it is depends on the majority of firms' predisposition to comply. But it considers only the relative costs of persuasion and sanctioning in the context of the relationship between the regulator and regulatee. As Kagan and Scholz argue, the costs to the agency in political terms may be far higher if it is shown to have been adopting an inappropriate persuasive approach than if it has been adopting an inappropriate

sanctioning approach (Kagan and Scholz 1984). Being a soft touch and 'letting firms get away with it' will receive far greater social, media and political approbation than being tough. Organisational preservation, particularly if the agency is in the spotlight, may thus favour a 'follow the rules' strategy.

The pyramid also remains trapped in the compliance / deterrence dialectic. Of Kagan and Scholz's potential enforcement strategies, it considers only regulators as politicians and regulators as policemen: strategies of persuasion or punishment. It says that in enforcing regulation regulators need not be either one or the other, they can be both. It does not say that they can and should be something else. That 'something else' could be conciliators, it could be educators. Keeping within Kagan and Scholz's categories for the moment, it ignores the fact that persuasion might not work, not because the firm is amoral, but because it is incompetent. Parker, for example, has argued in effect that regulators should pay more attention to being educators, as well as punishers and persuaders (Parker 1997a and forthcoming). They should focus on building regulatory capacity within organisations, and on 'meta evaluation' of corporate compliance efforts.

There are also other strategies that could be adopted that focus more strongly and systematically on providing 'carrots' for compliance, as well as sticks. Such strategies could include 'risk based enforcement' methods, for example, in which greatest enforcement efforts are focused on those firms posing the highest risk. This could be relatively informal, assessing firms with good compliance records less frequently than those with poor ones, for example, or it could attempt to be more formalised. An example of the latter is the UK Financial Services Authority's proposed approach for enforcement activity based on a risk assessment process which scores the risk of non-compliance against the impact of the risk should it occur multiplied by the probability of it occurring. The impact of non-compliance is to be gauged by the systemic nature of the firm, its perceived importance and so impact on market confidence, the number and nature of retail customers affected and whether compensation is available. The probability of failure is gauged by the quality of the firm's internal controls, its business strategy, its ability to absorb market volatility, and its customer relationships (likelihood of misselling). The risk assessment process will result in a classification of firms and in variations in intensity of the FSA's supervisory relationship with firms. Those with a high grading will receive intensive monitoring, those with a low grading will be assessed on the basis of self reporting combined with visits as necessary in relation to specific issues (for example those indicated as areas of concern in a product risk assessment) (FSA 2000). The scheme is only expected to become operational in April 2001, and so it is as yet too early to assess its effectiveness. Certainly the initial description of the process suggested many details were still to be worked out (for example the relationship between firm specific risk and product risk in monitoring, the relative weights of impact and probability in the assessment process, how the qualitative and to an extent intuitive assessments will be converted in to quantitative measurements and how the rating will operate in practice). The indication is also that resources will be focused on the larger firms, though research in other areas suggests that greater results can be achieved if resources are focused on mid-range firms (Gray and Scholz 1991). But it is an experiment that will bear close monitoring.

Other ways in which regulators could respond to the firm's efforts, or lack of effort, to comply could include taking the firms' past compliance record into account in determining the sanction imposed, or in determining whether there is liability at all. This is often informally manifested in the construction of compliance: firms with good past records of compliance are more likely to have a one-off breach (eg pollution discharge) treated as an accident that in normal circumstances does not merit the imposition of a sanction than firms with a poor past record. But leaving such 'carrots' to the discretion of the enforcement officer without any formal recognition in law leaves the officer and agency open to charges of discrimination and illegality (for detail see Parker, forthcoming).

Better therefore to have such carrots legally sanctioned: to allow risk based monitoring, for example, or to build in 'due diligence' defences, or to take the implementation of effective internal controls into account in determining the sanction to be imposed (see also Parker, forthcoming).

Finally, attention should be given to other aspects of enforcement practice. These include the balance that should be struck between proactive and reactive enforcement methods. Should the agency operate principally by responding to complaints or should it take a proactive approach to enforcement? If the latter, what types of investigations or inspections should be conducted (forewarned, spot checks)? What should they focus on (records, practices, internal compliance systems)? How should information from individual complaints or inspections be collated to see if there is a pattern of non-compliance emerging which is systemic in the industry and which might require a more systematic and focused response from the agency at a policy level? How could agencies gain information about a firm's compliance other than through the firm itself - what role is there, if any, for external verifications, 'mystery shopping' and the like.

In sum, whilst much academic and policy attention has been given to the development of general regulatory strategies: the relative merits and demerits of different forms of self regulation, voluntarism, economic instruments, information instruments and their various combinations (see eg Gunningham and Grabovsky 1998), because of the preoccupation with the punish - persuade dichotomy, far less attention has been given to the development of strategies for enforcement, either outside or within the command and control model in which many regulatory agencies currently operate, or to the relationship between enforcement strategies and general regulatory design. The enforcement pyramid managed to go somewhat to breaking the punishment / persuasion mould, but is constructed from the same basic elements. Rather than focus only on punishment and persuasion, we should think in terms of education, conciliation and compensation. Rather than the lexical ordering of the pyramid, we should think in terms regulators having a range of tools from which to fashion a targeted response. Rather than focus only on whether or not to sanction, we should focus on how best to deploy limited resources: whether to be proactive or reactive, what and how to inspect, how often, what to look for, how to interpret what is found.

Regulating regulators, 'managing' discretion

How can we try to ensure that enforcement officials operate in the way wanted? We have to start by building on the understandings of first, the nature of the enforcement task in any one regulatory system, and second, the nature of the factors influencing how enforcement officers currently operate, and tailor our proposals accordingly. This is likely to mean that one blueprint will not suit all types of regulatory and enforcement situations. Strategies to regulate the conduct of a few dozen case officers working in a utility regulator ensuring that operators comply with their licences and do not act anti-competitively are unlikely to be appropriate for regulating a few hundred local trading standards officers responsible for enforcing the requirements of several regulatory regimes, and vice versa. At the very least, the inescapable operational autonomy of the latter is likely to be far higher than the former. Nevertheless, some general suggestions can be made, both of which strategies might work, and which are unlikely to do so.

Altering enforcement behaviour requires attention to be paid to those factors that influence current practices. The discussion above identified four groups of factors: the legal framework, the relationship between the regulator and regulatee, organisational practices, and the broader political, economic, social and moral context. Reformers can do little or nothing about the last, and can influence the second (relationships between regulator and regulatee) only indirectly. Attention will

thus be paid to the legal framework and to organisational practices. We should not necessarily have to cast reforms in the mould of the present, however, and so attention will also be given to alternative strategies. In broad terms, the proposal is that the regulation of the internal operations of an organisation is similar in many essential respects to regulation of one organisation by another. We can therefore draw on alternative strategies of regulation proposed for governmental regulation of firms to develop proposals for governmental regulation of government.

Adjusting the legal framework of regulation

The legal framework of the regulation is not primarily directed at those who will be enforcing it, but nonetheless it can affect enforcement strategy used. The legal framework is the enforcement officer's basic toolkit. It is neither exhaustive nor determinative: officers use and devise other tools, and may not use those tools in the way that those forming the rules (legislators, other department or regulatory officials) may want, but it provides the basic framework that officers have to work within, or work around. Three aspects of that framework can affect enforcement strategies: the type of rules that are used, the range of sanctions that are available, and the appeal or review mechanisms that are put in place.

The type of rules

Rules may be characterised as having four dimensions: substance, character (eg be mandatory or permissive, or be target standards, performance standards, specification standards), status and sanction attached (legal, quasi-legal, non-legal, criminal, civil, administrative, voluntary), and structure (precise or vague, clear or opaque, simple or complex) (J. Black 1997: 20-30). Each dimension represents a decision point for a rule maker: what should the policy be that the rule is trying to further, how should it be expressed, what status should the rule have and what sanction should attach to it, and should it be vague and simple (a 'principle', or 'standard' or 'general rule'), or should it be precise and simple (a 'bright line rule') or precise and complex (a detailed rule). Whether the rule is clear or opaque is less under the control of the rule maker: a clear rule is one which most of those reading the rule would interpret in the same way. Clarity is thus a function of the interpretation a rule receives in a particular interpretive community (ibid, 24). Each type of rule has its own advantages and disadvantages, and therefore trade offs have to be made at each decision point. For example, a general rule which states that polluting emissions should be 'the lowest levels that are reasonably practicable', or that utility firms should not abuse a dominant position or unduly discriminate in the provision of services, leaves considerable discretion to the enforcement officer to determine the standard. It thus facilitates (mandates?) a negotiative approach. This gives flexibility, but raises concerns of consistency and certainty, and leaves the enforcement officer vulnerable to challenge by a firm that does not accept the officer's or agency's interpretation. In contrast, both enforcer and firm have less discretion, or room to exercise judgement. If the rule states for example that emissions must be restricted to x parts per million, or that prices of specific products should not increase by more than 2% a year, though at some loss of flexibility. We should be careful about being too categorical in these generalisations: as noted above, one should not presume that the precise rule leaves no discretion, nor that what is accepted on the ground as compliance will conform to that standard (ie may be higher or lower), but precise rules leave less room for judgement than a more general rule.

The implications of rule type for enforcement strategies has been discussed by Baldwin. He argues that different rule types inhibit or facilitate different types of enforcement strategy (Baldwin 1990, 1995). General can facilitate an educative approach to enforcement, whereas such an approach can be hindered by a mass of detailed rules which may simply serve to alienate the regulated. But, as

noted above, different enforcement approaches may be appropriate in different contexts. With respect to political citizens (or in Baldwin's terms the well intentioned and well informed), rules play little role in compliance. Rather they are used to keep the firm informed and provide an agreed agenda for discussions and promptings by the regulator. The rules thus have to be accessible, and in a form that corresponds to the firm's capacity to absorb them. A large body of precise rules can be difficult to absorb and alienating: ignorance and the sense of irrelevance could then hinder negotiation and compliance. General rules with user-friendly guidance manuals are therefore preferable. Similarly, for the organisationally incompetent (Baldwin's well intentioned but ill informed) a series of general rules with user-friendly 'how to comply' guides are necessary as the enforcer needs to give advice and information and adopt techniques of promotion or education.

With respect to the irrational non-compliers, however, rules will not produce compliance and persuasion has to be backed with realistic threats. Precise rules tend to be more effective as they enable the enforcer to show that the rule had been breached, without having to engage in a debate about whether something is or is not 'suitable' or 'appropriate', or what others do and how they have been treated (see also Hawkins 1984). Detailed rules also offer assurances of consistency and so enhance persuasion/negotiation in enforcement.

The most problematic category are the amoral calculators. Baldwin did not consider this group, but neither detailed rules nor general rules are on their own likely to be successful in dealing with them. Detailed rules are vulnerable to strategies of 'creative compliance': compliance with the letter but not the spirit of the rule (McBarnet and Whelan 1991). General rules are vulnerable to challenges as to their interpretation and application. Amoral calculators are likely to contest the agency's interpretation of the rule and assessment of compliance. The structure of appeal and review mechanisms then becomes highly relevant for determining who controls the interpretation of the rule, and thus the agency's authority, *de facto* and *de lege*, for insisting that its interpretation is that which is 'correct'.

Baldwin's proposed solution is that rule makers, be they legislators or agencies, should consider in advance of making a rule whether regulatees are likely to be political citizens, amoral calculators, organisationally incompetent or irrational non-compliers. They should determine which type of regulatee is likely to be predominate, and should then fashion rules to the enforcement strategy that will be appropriate for that type of regulatee (Baldwin 1995: 174). I agree that it is critical that rule makers take enforcement approaches into account when forming rules, preferably including enforcers in the rule making process. But given that it is unlikely that any one strategy will be appropriate for any one regulatory system, a better response is to assume that enforcers will need to be able to deploy all strategies. They should therefore be provided with as wide a range of ways of communicating the required conduct as possible.

Rather than opting for one rule type, rule makers should adopt a tiered approach to rule design in which rule types are combined in such a way that each tries to compensate for the limitations of the other. This will require legislators (and parliamentary draftsmen) to be sensitive to issues of rule design, and to confer power on regulators to make rules of differing legal status. For example, general standards of legal status (in statute or made by regulators) could be backed by detailed guidance which has no legal status, or has evidential status, or which creates safe harbours. The general, purposive rule can forestall a creative compliance approach, but is likely to be over inclusive and give rise to problems of consistency and certainty. These problems can be addressed by having more specific guidance on what constitutes compliance, the general rule serving to block any loopholes that the guidance might create (see further Black 1999). General rules could alternatively or in addition be elaborated by rules written by firms themselves and approved by regulators (Ayres and Braithwaite 1992), and be accompanied by 'how to comply'

guides (see further Black 1999). In all cases, efforts should be made to use language that regulatees understand. This is a call for 'plain English' in part, but also for the appropriate use of technical or other terms which carry particular but widely recognised meaning for regulatees. Legislation and regulatory rules should be drafted primarily with those that they are intended to regulate in mind, not the courts.

Range of sanctions

The impact of the types of sanctions that enforcement officers have on the enforcement strategy that they can use was noted above. As discussed above, officers should be given a wide range of sanctions so that they are able to tailor the sanction both to the offence and to the offender. What type of sanctions should be available is a separate question, and is discussed by other papers. It is worth noting two points briefly here, however. First, that enforcement officers' attitudes to what type of conduct merits prosecution (it may not apply in the case of non criminal sanctions) injects elements of culpability and criminal morality even if in law the offence is a strict liability offence. Secondly, that the debate of the type of sanction should not just occur as to whether the sanctions should be imposed by a court, tribunal or regulator, or whether they should be civil or criminal, but should include consideration of sanctions that involve the firm in its own punishment in such a way that is likely to prompt the firm to institutionalise systems of compliance in the future, not simply pay out fines for misdemeanours in the past (see Ayres and Braithwaite 1992, 41-44; Parker forthcoming).

Appeal and review mechanisms

The third aspect of the legal framework that can affect how the agency operates is the nature and structure of mechanisms for the appeal or review of its decisions. In discussing some agencies' reluctance to seek prosecution, the feeling that the court would adopt an interpretation of the requirements that the agency considered to be too lax was cited as one reason why agencies can be reluctant to prosecute. Agencies are likely to prefer to use administrative sanctions, partly because they are quicker and cheaper to use, but also because the enforcement officer or enforcement team may believe that its interpretation of the rule will be shared by the part of the agency or separate body that is responsible for discipline and enforcement.

This cosy world of shared interpretations and commonly imposed standards amongst different parts of the regulatory system can be upset by mechanisms of review and appeal to other agencies, tribunals or courts: that, after all, is their function. Those mechanisms can play an important role in ensuring the accountability and legitimacy of the administrative body. The extent to which they are effective in achieving that aim can be disputed: the impact of judicial review of agency decisions, for example, is far from clear cut (see Richardson and Sunkin 1996; Loveland 1995, Halliday 2000). The point to note here is the significance of appeal or review mechanisms for the control of the interpretive framework of the regulation, and thus for questions of rule design and for enforcement practices. General rules, as noted, facilitate a persuasive and educative approach, but leave the regulator vulnerable to challenge over its interpretation and application of the rule. Detailed rules can forestall such challenges, or if they do not they can perhaps go some way to ensuring that if such challenges are made that the body that is reviewing the agency's decision will adopt the same interpretation as the regulator. Detailed rules however bring their own problems, as noted. The solution of a tiered approach to rule making could go some way to addressing the issue of interpretive control, but the problem with only specifying the meaning of the general rule in guidance is that it still leaves room for interpretive manoeuvre. That is the advantage of such a strategy in the context of the regulator-regulatee relationship; but its disadvantage in the context of the regulator-reviewer relationship. But it is important to the development of regulatory interpretive

communities that the regulator has some form of interpretive authority, and the development of interpretive communities is important to ensuring the effective implementation of regulation. The route which perhaps best balances the competing concerns of accountability of the regulator and allowing the regulator sufficient interpretive authority is for the regulator's interpretation of the rules to be open to review on the grounds of rationality, not appeal on a point of law (see further J. Black 1998b; and for a highly critical account of the effect of appeal on a point of law on the operation of the UK pensions ombudsman see Nobles 2000).

Organisational culture and practices

Organisational factors were noted above to be an important in influencing enforcement conduct, both directly, and through affecting the ability of enforcement officials to develop particular kinds of relationships with regulatees. Thus it was suggested that those agencies adopting reactive enforcement strategies (responding to complaints, or accidents) are perhaps more likely to 'go by the book' than those that adopt proactive strategies; that frequent rotation of personnel can increase the relational distance between individual officers and firms (though relational distance may remain low between senior managers of firms and agencies), and that internal pay and promotion measures can affect at least how officers present their work. Regulation of enforcement officers could thus be based in part on adjustments in these aspects of agency operations. Overall, however, it is organisational culture which has the greatest impact; trying to influence that, however, is the hardest task.

Developing Enforcement Principles: changing practices through rules

One of the standard responses for how to regulate bureaucrats is to have rules (Davis, 1969). Rules governing officials are meant to serve two principal functions: to structure and constrain discretion, and to provide benchmarks for accountability. The UK government has recently adopted such a rule based approach, developing principles that should guide enforcement officers in conducting their enforcement duties. The Enforcement Concordat (March 1998) sets out six principles of good enforcement policy: performance of officers to be measured against published standards, openness in dealing with businesses and others, helpfulness, courtesy and efficiency in advising on legal requirements, publicised complaints procedures, proportionality, and consistency in the standards of conduct required. These are supplemented by principles of good enforcement procedures. First, that advice will be put clearly and simply and confirmed in writing, and legal requirements distinguished from best practice advice. Second, that before formal enforcement action is taken enforcement officers will provide an opportunity to discuss the circumstances of the case and resolve points of difference unless immediate action is necessary (for example in the interests of health or safety or the protection of the environment). Thirdly, where immediate action is necessary, an explanation of why such action was required to be given at the time and confirmed in writing within 10 working days. Finally, to give advice on rights of appeal in writing at the time the enforcement action is taken.

The Concordat has been adopted by the majority of local authorities and by over forty central government agencies, including the Health and Safety Executive and the Inland Revenue. In addition, the ministers are given powers in the Regulatory Reform Act 2001 to draw up Codes of Practice for enforcement practice where they consider it necessary to establish fairness, transparency and consistency. The codes do not impose a legal obligation, but failure to comply with them may be taken into account by court or tribunal if formal enforcement action is taken by

the relevant regulatory body (Regulatory Reform Act, s.9). It is envisaged that the power will only be used, however, if the Enforcement Concordat has not been widely adopted (CAB 197/99).

The Enforcement Concordat is just one example of rules or principles which are meant to guide, control, manage or regulate the ways in which government officials act. The presence, and sheer abundance, of internal administrative rules has been well documented (Baldwin and Houghton 1986; Baldwin, 1995). Many of these are specific to those responsible for the delivery of specific services (eg the health service, schools, prisons), but also some centrally imposed on all government departments and other public bodies. In the UK, the Modernising Government initiative has spawned dozens of 'how to' guides to improve public sector performance. These include guidance on conducting regulatory impact assessments (Cabinet Office, RIU 2000; Froud et al 1998), improving local authority services (Cabinet Office, Service First 1999). In Northern Ireland, 'mainstreaming' requirements have been introduced to implement the requirement of all public bodies to ensure equality in public sector service provision (McCrudden 1999, 2001; Equality Commission of Northern Ireland 2000). Some of these rules or codes, like the Enforcement Concordat, are very general in nature; others, like the Regulatory Impact Assessments (RIAs) are more specific. Some, like the mainstreaming requirements for ensuring equality in public service provision in Northern Ireland, are legislatively imposed, others, such as the RIAs and Concordat, are administrative requirements only.

Such statements of principle have a role to play. At the least, they are constitutional wish-lists: statements of the values that it is expected should be pursued and the standards of conduct required, and provide benchmarks for accountability. Exactly what the content of a statement of enforcement principles should be will depend on the political and moral context in which they operate, and the preoccupations and concerns of the polity of the time. One suggested set of principles would be that enforcement should occur in accordance with the following standards

- Legality
- Consistency
- Proportionality
- Non-discrimination
- Due process
- Transparency and accessibility
- Efficiency and expertise

There are many other candidates for inclusion (effectiveness? 'fairness?' 'helpfulness?'), and there are tradeoffs to be made between them (see eg Hood, Baldwin and Rothstein, 2000). But how effective are such principles likely to be? The answer is that on their own, they are unlikely to have much impact. This is often in part for the usual reasons of institutional resistance to change, or even objection to the goals of the internal regulatory system - bureaucrats, like firms, can be amoral calculators or irrational non-compliers (for an example of bureaucratic resistance to compliance costing of regulations see Froud et al 1998). But even if all are signed up to the goals of the internal regulation (or at least, like many firms or individuals within them, the goals lie within the individual's 'zone of tolerance' (Selznick 1992) such that they are likely to be inclined to comply unless the course of conduct required is in serious conflict with their own personal morality) inadequate compliance is likely to result.

This is because of the problem of 'honest perplexity': wanting to comply with the rule, but being totally unsure of what type of conduct will be considered to be compliance (J. Black 1997). Standards of 'consistency' or 'proportionality' may be unobjectionable goals, but simply setting

them out is of little assistance to enforcement officers on the ground in how to handle a day's workload. Like well intentioned firms, enforcement officers need advice, guidance and education as to the meaning of the rules or principles they are being subjected to. General principles or rules which require 'transparency', 'fairness', 'consistency', 'proportionality' require the exercise of judgement, but they cannot of themselves create informed judgement as to their interpretation and application where none exists. General rules or principles on their own thus run the risk of not changing conduct at all, either because the regulated are unaware of how it should be changed, or because they assume that what they are already doing is in compliance with the rule. They simply use their own existing set of norms and understandings when reading it, with the result that the rule does not actually affect those norms or understandings. In other words, unless the person reading the rule knows what it requires, they will either not understand it or adopt an understanding which is not that which the regulator meant. This may seem back to front; how will the person know what to do otherwise than by reading the rule? But it is the understandings which inform the rule, not the other way around (J. Black 1997, 1999).

General rules or principles on their own will thus result in standards of compliance which are either low overall, or which are patchy, with some enforcers (like firms) exhibiting better practices than others. One common way of trying to guard against inadequate compliance is again more rules (eg Diver 1983). But simply increasing precision will not work, for all the reasons it does not work when detailed rules are used to regulate firms. For not only can rules, as noted above, never fully control discretion, they bring their own problems. These included problems of '*rule overload*' (too many rules for the addressee to be able to absorb and remember, which means they are unlikely to comply with all of them), *rule system complexity* as detailed provisions create internal inconsistencies and contradictions, *creating loopholes* and facilitating '*creative compliance*'. In a bureaucratic context the latter is likely to manifest itself in formalism and inflexibility: following procedures where these impede, or at least do not assist, the achievement of the substantive result.

Certainty as to what the general principles of an enforcement code actually mean in practice comes not through detailed specification, but through shared understandings. Whether a rule or principle is certain depends on whether the terms in which it is expressed are commonly understood by those applying the rule. A rule or principle is clear, and thus certain, if the interpretive assumptions and procedures are so widely shared in a community that the rule bears the same meaning for all. The greater the shared understanding of the rule and the practices it is addressing, the more the rule maker can rely on tacit understandings as to the aim of the rule and context in which it operates, the less the need for explicitness and the greater the degree to which simple, vague rules can be used. So in these circumstances, general rules can be certain. The next question, then, is what is necessary to create those circumstances, and how realistic is it to suppose that they can be created at all?

The essential prerequisite for certainty is mutuality of understanding and thus of interpretation. In ensuring this mutuality all those who will be involved in the rule's application and interpretation have to be considered: enforcers, line supervisors, adjudicators, regulatees. Each has to share the others' interpretative approach, and each therefore needs to be aware of that of the others, and make the others aware of its own. This mutuality can only partly be ensured through rules. Instead, we have to step outside the rule book, and look directly to communications which are external to it: to conversations as to the meaning and interpretation of rules, as to what they require and when they apply (Black 1997, 1998a).

So developing principles to guide enforcement conduct is a start, but it is only that. Principles that require enforcement conduct to be procedurally fair, consistent, non-discriminatory, proportionate, and transparent, for example, can be relatively easily devised. But in determining what it is that

those principles mean, an interpretive community has to be developed through dialogue between regulators and enforcers at all levels - from senior directors down to field level officers and back again, and between regulators, enforcers, the regulated, and other stakeholders.

The emphasis on conversations is deliberate: it indicates that the process of forming interpretive communities cannot be a monologue engaged in by the regulator (be that the legislator, minister, or senior officials external or internal to the agency) and listened to by those whose behaviour the rules are meant to govern (in this context, enforcement officers). It is important to stress that the process should not be a top-down exercise: executives handing down rules for field officers to follow. If anything it should be the other way around. Nor should the process be confined to regulators: the regulated and other stakeholders should also be involved through the formation, for example, of forums, 'partnerships' or other mechanisms (for examples at a local authority level in the UK see the Best Practice Database (www.goodpractice.org.uk). For there will inevitably be tensions and tradeoffs in any list of principles, those need to be recognised and a view taken on how they should be addressed. But such an exercise is an important step in rendering operational the exhortations for best practice that the principles express, and to internalising those principles in the workings of the enforcement agency. It is a topic to which we will return below.

Again the issue of interpretive control is critical. It may be felt that to ensure consistency of approach, for example, and to monitor the development by agencies of codes of practice for elaborating and implementing the enforcement principles that those codes of practice should be approved by a central body: Ayres and Braithwaite's model of enforced self regulation applied within the bureaucratic context. This would not be unusual in the context of central executive control over agencies: in the UK all RIAs have to be assessed by the Better Regulation Taskforce, for example, and the role of the Office of Management and Budget in using review of rules and regulatory impact statements to control federal agencies has been long noted. But as observations of the OMB have noted, such centralised control has to be carefully managed if the advantages of self-written codes are to be captured (responsiveness, appropriateness, regulatee and other stakeholder input, sense of ownership and so on). Over-insistence by the approving body that the codes comply to its own model can result in inappropriate and excessive detail, and serve to achieve the opposite of what was intended. Any mechanism of central approval, therefore, must allow sufficient margin of respect to those formulating the codes under the overall umbrella of the principles.

Changing practices by exerting other means of control

Rules, then, have a role, but they also have their limits. Just as there is a range of alternative techniques for regulating those outside government, however, there is a range of techniques for regulating those inside government, or at least those who are operating on its behalf. Those strategies are often grouped into four or five categories: legal instruments, economic instruments, information, some form of self regulation (eg Gunningham and Grabovsky 1998), to which some have added technologies (Lessig 1998; Black 2001 forthcoming). Others group the tools of governance into state, market, community, associations and networks (eg Rhodes 1997). In his study of 'the art of the state' Hood identified four potential control strategies that could be used to regulate the state, the first three of which have broad parallels in the strategies used to regulate businesses: oversight, competition, mutuality, and contrived randomness (Hood, 1996, 1998; Hood et al 2000).

Hood et al's study of regulation inside government found that oversight, mutuality, competition and contrived randomness can each can be an important control strategy in different areas of regulation

of government, either alone or more usually in combination with another. Contrived randomness was used least, and most control forms usually involved some element of oversight. (Hood et al, 2000). Competition and mutuality in particular were found to be closely related, indeed sometimes hard to distinguish. For example, the system for assessing the performance of UK universities was in places classified as one based on mutuality (ibid: 17; 45) and in others as one based on competition (ibid: 51, 177).

Hood's typology is a useful heuristic device, though there is less difference between the control strategies he identified and other typologies than might at first appear. 'Oversight' uses a form of the traditional 'command and control' techniques: rules, with compliance overseen by hierarchical superiors or inspectors, and sometimes backed by formal or informal sanctions. It is thus analogous to the use of legal instruments backed by sanctions, and so to some traditional understandings of 'regulation'. Control through competition in Hood's taxonomy is control through rivalry and choice (eg parents' selection of schools, patients choice of hospitals). It is thus a form of market based techniques of control, which may additionally include the use of economic incentives and information. Mutuality is control through peer pressure and peer assessment: ie institutional norms, which may or may not be formalised, of what is the 'right' thing to do. Mutuality is thus a form of self regulation or regulation through community (or possibly association). 'Contrived randomness' is the strategy which has no obvious parallel in the regulatory strategies that have been elaborated for organisations outside government. Contrived randomness is control through unpredictable processes or payoffs, for example posting personnel unpredictably in different locations to prevent corruption. On the other hand, the role of networks in regulating those inside government does not form part of Hood's analytical matrix, whereas it has been suggested as a technique of governance or regulation outside government (eg Rose 1999; Rhodes 1997). That is not to say that randomness does not or could not play a role in regulation outside government and networks could not or do not play a role in regulation inside government; it is rather to comment that the typologies do not match in all respects.

Hood's typology nonetheless provides a useful framework in which to consider different forms of bureaucratic regulation that could be adopted, either alone or in combination, and examples of each being used in practice can be found. Oversight is probably the most common: line reporting up a hierarchy of officials based on standards. As noted, it is the bureaucratic equivalent of 'command and control' regulation: stipulation of standards to be attained with some form of sanction attaching for their breach. The standards may be formal or informal, and sanction could range from a cut in agency resources to lack of promotion of individuals, or take a more nebulous form.

Competition is increasingly important, and it is useful perhaps to distinguish between internal and external competition. Internally driven competition may take the form of performance related pay, or publicising relative performance figures. They could take the form, for example, of league tables comparing performance of individuals, divisions or different agencies against criteria of which officers or divisions have issued the most enforcement notices, or taken the most prosecutions (for an example in practice see Hutter 1997). Externally driven competition could take the form of firms choosing which regulatory agencies or inspection teams they want to be inspected by, or which tax offices they preferred to deal with, for example, though the dangers of inefficient regulatory capture from such strategies could be too great to merit their adoption.

Contrived randomness as a control strategy is used by enforcement officials with respect to firms in the form of unannounced inspections, but could also be used with respect to officials themselves: superiors or other assessors of performance could perform spot checks by coming unannounced on inspection visits. Or contrived randomness could take the form of frequently rotating personnel. This may indirectly influence the regulatory approach adopted in that it increases the relational

distance between individual firms and individual enforcement officers. However, as noted above, this may be negated if more senior executives within the agency have a close relationship with the firm, and those executives can either override the decisions of lower level officials, and / or officials act in such a way as they know would receive the approval of senior management.

Finally, mutuality, that is peer pressure and peer review, as we have seen, is in practice a strong regulator, with institutional norms of what counts as a 'good enforcement officer' significantly affecting enforcement practices. Existing systems of mutuality could be harnessed, and to an extent shaped in a number of ways. These could include, for example, self audit coupled with benchmarking exercises (eg the UK Public Sector Excellence Programme, assessed by Samuels 1999), through best practice learning, which could involve, for example, creating forums in which enforcers in different areas can exchange ideas, the creation of joint codes of enforcement (for examples see Service First 1999), creating coherent career structures for regulators, as has occurred for compliance officers in business (Parker, forthcoming) and restructuring organisations to bring disparate specialists together under broader umbrellas (suggested by Hood et al, 212).

These suggestions go far beyond the mere issuance of a set of principles, but to varying degrees still depend on standards of performance being agreed, and to varying extents specified. As such, therefore, they still run into the problem that they are likely to distort enforcement conduct as they replace qualitative judgements with quantitative measurements, or substantive rationality with formal rationality. It is the classic audit problem (see Power 1997). Strategies that rely on audit or league tables of performance figures in the exercise of control can have unintended, and disadvantageous, consequences. They can disempower those who do and simply empower those who count, and they skew behaviour towards those aspects of the work which will be measured to the detriment of other aspects of the task which are equally if not more important, but not as easily measurable. Their distorting effect is magnified if they are used as the basis for control, as in strategies based on oversight and competition.

This is a particular issue for enforcement practices as the normative model of how enforcement officers should act may not be reflected in the criteria against which they are assessed. The traditional ways of measuring performance are to look at the number of notices issued, prosecutions brought. In other words, it is to look at the rate at which sanctions are deployed. This is, after all, the most visible evidence of enforcement activity. This might be appropriate if the most appropriate task for enforcement officials was to issue sanctions. However, as emphasised above, the sanctioning approach is neither that which is adopted in practice, nor that which normatively should always be adopted. Instead, the preferred enforcement approach, normatively speaking, is one in which officials use a combination of sanction, suasion and education, varying with the nature of the offence and the conduct of the regulatee. Performance standards that do not reflect this approach, whether written in a code of practice or applying internally in the form of criteria for promotion or salaries, are simply inappropriate. Such enforcement approaches need systems of control and accountability based on criteria relevant to those approaches, not to an approach which is quite distinct, although easier to measure. To this extent, the inclusion in the Enforcement Concordat of standards that emphasise advice and education, if inelegantly framed, reflects the nature of many enforcement officials' work far more accurately than performance criteria based on a sanctioning approach.

Devising appropriate performance criteria is only one part of the task; the other is obtaining information on which to base a performance assessment. External audits, individual self assessments and traditional line reporting and managerial oversight are well recognised techniques. But they are not enough. The information on which assessments are made should come from as many sources as possible. In management terms, there should be '360 degrees' accountability. That

is, accountability should not just be either upwards, or downwards, or laterally, but all round: all those who work with the individual should be involved in assessments of their performance. For example, how clear, accessible and valuable the compliance advice was that was given is something that only the firm that has received it can assess. Some local authorities in the UK are beginning to experiment with ways of getting information from firms as to their views on enforcement officers' activities. One example is to ask businesses to fill in questionnaires after inspections and audits on food safety and health and safety which include questions about the nature and quality of information received from staff (Service First 1999), the responses to which should be collated by a someone other than the individual inspector.

Internalising standards

All the methods outlined above are forms of external control. But as many studies of firms' responses to regulation have pointed out, regulation is only fully effective if it is internalised. That is, if it becomes part of the internal morality of the organisation: if it is institutionalised (Selznick 1992; W. Scott 1994). And as we have seen in the discussion above, one of the most important influences on how enforcement officers act is organisational culture. The formal systems of control, as we have seen, are always supplemented by informal structures: practices, attitudes, experiences, personal codes of morality. The task for any system of regulation, be it of firms, bureaucrats, employees, etc, is to ensure that those informal systems support the formal system by enhancing cohesion, initiative and morale (Selznick 1992: 235).

Here we can link back to the discussion on interpretive communities. Building an interpretive community is not just about determining the technical meaning of terms: interpretive communities are not simply lexicographical communities. It is about developing shared understandings of the goals and values of the regulatory system which may only be partly, and often inadequately, expressed in rules, and for developing ways for addressing the inevitable conflicts, inconsistencies and trade offs that there will be between those goals. Of itself, the development of an interpretive community does not mean regulation has been institutionalised, but building interpretive communities is part of the institutionalisation process (J. Black 1997a; Meidinger 1987; Parker 1999).

Hood et al's work on regulation inside government in the UK did not address the need to internalise regulation, or how ask that need could be met (Hood et al, 2000). Instead, we can again learn from research on business responses to regulation, particularly from the work of Christine Parker (Parker, forthcoming). She suggests that for regulation to be successfully internalised, three stages have to be accomplished. First, the organisation has to be prompted to commit to a set of principles, usually through threats of enforcement action if they do not. Second, the acquisition of the skills and knowledge required to develop systems that ensure compliance has to be facilitated. Third, and most problematic of all the three stages, there has to be institutionalisation of purpose.

These stages could be adapted to regulatory bodies and enforcement agencies. The first stage, prompting commitment, could take the form of requiring different regulatory bodies or enforcement officers to devise their own codes of practice as outlined above, and to devise and accountability or feedback systems, elaborating on the central enforcement principles. This should be done in consultation with regulated firms and other interested parties, with the threat, for example, that if they do not a code will be imposed upon them (assuming that that will be perceived as a threat and not just a way to avoid having to devise a code), and / or that funding will be reduced until a code is in place (assuming the agency is funded by government). Examples of this occurring in practice are the development of 'partnerships' or business forums between local authority enforcement officers and local businesses.

The second stage, helping agencies to acquire skills and knowledge to develop systems could be achieved through some of the strategies of mutuality outlined above. These include benchmarking, and providing support to agencies who are attempting to evaluate their own performance as part of a benchmarking exercise, and disseminating examples of best practice. The mechanisms for dissemination may need to vary between different participants in organisations. Some may be happy using web databases of best practice examples, others may prefer personal interactions. Thus a range of mechanisms should be used including the creation of networks and forums in which information can be exchanged and experiences compared, web databases, seminars or other meetings. It could involve the identification of 'beacons' or 'champions' of best practice and the identification of facilitators who can act as co-ordinators of information and knowledge transfer (see eg Office of Public Management 2000). Simply providing information is not enough, however, just as simply providing information to businesses on how to comply is not enough. What is needed, as in regulation of business, is the development of a community of meaning and a community of commitment. Those involved need to trust and respect the source of the information, have a shared understanding of the language used and the aims sought, see its relevance to them, and the benefit of acting on it (see also *ibid*).

This brings us to the third stage, institutionalisation of purpose. At the level of process this involves two things: self audit and reporting, and external audit. Self audit should take the form of a self evaluation of the operation and effectiveness of the enforcement systems put in place, and reporting on that self evaluation. That report should indicate what basic principles of enforcement practice are in place, and should focus both on substantive outcomes as well as procedures followed. As noted above, evaluation of substantive outcomes is likely to be the most difficult to achieve if that evaluation is to fit with the systems of enforcement that are desired, for the conduct which is most visible: sanctioning, is conduct which forms only a small part of enforcement agency work. Part of the second and third stages thus needs to involve designing criteria for the self evaluations that better reflect the work that agencies do. What happens to those reports is a complicated issue. Ideally for they should be made public for those interested in the agencies' work, including regulated firms, the public, ministers, legislative committees etc, to evaluate them and to call them to account. In practice, knowing that they are to be published is likely to lead to them being less than frank (Power 1997). But the point of the self audit is to get the organisation to recognise its failings in order to correct them. It may be that in order to achieve that goal, only a relatively small group should see the report, which should include some form of stakeholder representation, the cost to such a strategy being that only a sanitised version is produced for external consumption. A solution that is not totally satisfactory, but perhaps the best that can be obtained.

There is some evidence that work under the Enforcement Concordat is proceeding, in form at least, at all three stages. With respect to the first stage, the voluntary code has been widely adopted, and is backed by the threat that one will be imposed if the agency has not adopted the code voluntarily, and a further 'stick' used to ensure compliance is that failure to comply with a code will be taken into account should legal proceedings ever be taken by the authority (though whether this has an impact is likely to depend on how critical the authority sees legal action to be to its enforcement policy both in general and in the particular case where the code has been breached). With respect to stage two, there has been some work done by the Service First team and by, for example, the Convention of Scottish Local Authorities, the Local Authorities Co-ordinating Body on Food and Trading Standards on developing arrangements for liaison between local authority enforcement bodies and for disseminating best practice examples. Also, as noted above, benchmarking projects are being used with respect to service delivery broadly, though there have been considerable difficulties in developing criteria to measure performance, and it can take some time to educate personnel within authorities on how to carry out self assessment procedures. Finally, enforcement

agencies are encouraged to use an Audit Proforma to assess their own performances in enforcement, and to use the Excellence EFQM Model (European Foundation for Quality Management, available at www.cabinet-office.gov.uk/eeg/2001/02.htm) to assess their performance overall as part of the Public Sector Excellence Programme. A preliminary assessment of the programme found that whilst government bodies compared favourably on many measurements to private sector bodies, there was considerable difficulty in devising adequate performance measurements for key aspects of public sector activity, and that it was important that resources were committed to educating staff in how to complete the assessments (Samuels 1999).

A further example of an attempt to internalise standards, but too nascent to be able to draw any conclusions on its effectiveness, is the mainstreaming of the equality requirement imposed on all public bodies in Northern Ireland (McCrudden 1999, 2001). Each public authority is required in carrying out its functions to have 'due regard' to the need to promote equality of opportunity between certain individuals and groups (persons of different religious beliefs, political affiliations, racial groups, age, marital status, sexual orientation, between men and women, between persons with disability and persons without and between those with dependents and those without) (Northern Ireland Act 1998, s. 75). All public authorities are required to submit an equality scheme to the Equality Commission which must show how the authority proposes to fulfil the section 75 duty. The Act sets out the minimum contents of such a scheme, which must include (inter alia) arrangements for assessing compliance, for consultation on policies, for monitoring the impact of policies adopted, for publishing the results of assessments and monitoring, and for training staff. The schemes must also conform to any additional guidelines produced by the Equality Commission (and approved by the Secretary of State). The Commission is under a general duty to keep under review the effectiveness of the schemes, and if it receives a complaint that the authority is not complying with its equality scheme, it must investigate or give reasons for not investigating. Mainstreaming is moving into stage two in the process: equality schemes have been drafted and knowledge on best practice is being disseminated via such mediating institutions as the Committee for the Administration of Justice. There is still a long way to go. But again it is an experiment which will merit monitoring closely.

Conclusion

Enforcement officials are thus continually required to exercise discretion, and the factors that influence how they act include their own personal backgrounds and interpretation of their role, the nature of their interactions with those they regulate, organisational practices and culture, the legal framework, and the political, social and moral context in which they operate. Furthermore, the organisations of which they form a part are complex, and whilst enforcement officers may operate with one set of practices and culture, there may be competing organisational cultures and policies in other parts of the organisation.

In considering how enforcement officers should perform their functions, principles of legality, political morality, efficiency and effectiveness all play a role. Thus officers should act within the law, and in accordance with principles of rationality, proportionality, consistency, non-discrimination, transparency and due process. These principles will of course conflict at different times and in different contexts, and trade offs will have to be made.

In terms of efficiency and effectiveness, officials should aim to be as responsive to business culture as possible. This requires that the legal framework in which they operate be appropriately designed. That involves adopting a compliance based approach to rule making, a tiered approach to rule design, and a wide range of sanctions for breaches of the regulatory requirements to punish

recalcitrants, and which also require recalcitrants to take some form of remedial action. It also requires that they have a range of 'carrots' available to encourage the development of effective systems of internal compliance. Those carrots could take the form of due diligence defences and / or definitions of liability which take due diligence into account. Legal authorisation for the adoption of risk based enforcement methods may also be necessary to avoid charges of inconsistency and discrimination.

In deploying their powers, officials should also be prepared to use strategies of education and conciliation in addition to those of punishment and persuasion. In using those strategies, they should attempt to be responsive to the motivations and actions of the firm, though this may be easier said than done. Understanding motivations is not always a simple task, and moreover, firms' motivations may shift over time, and may not be easily consignable to one particular category. Given the internal complexity of corporate organisations, different parts of the firm, or the same parts of the firm at different times or even simultaneously, are likely to have different motivations. For a responsive strategy to work, it is also important that firms understand the enforcement policy that the regulator is proposing to adopt, and that they are clear what the nature of the supervisory relationship will be.

Agencies should also engage in periodic reviews of the effectiveness of their monitoring practices. Those reviews should include consideration of the balance between proactive and reactive enforcement methods, the type of inspections should be conducted (forewarned, spot checks) and what they should focus on (records, practices, internal compliance systems). Consideration should be given to the development and operation of targeted or risk based enforcement methods (or a rationalisation of the informal methods that may exist), the development of forums and other mechanisms for building shared understandings between regulators and the regulated as to the standards of conduct required in language that they can understand, and as to the enforcement policies and practices of the regulator.

In attempting to 'manage' enforcement discretion, reformers should be aware of the difficulties inherent in that task. Attention should be paid both to the legal framework and to organisational practices and internal cultures. There is a role for a statement of enforcement principles, but those principles will need to be supplemented by shared understandings and commitment to their requirements, understandings that have to be common to those promulgating the principles, those in the agencies whose conduct they are meant to regulate, and the wider regulatory community. Those understandings have to be developed through conversations and consultations, and may ultimately be formalised in codes of practice formulated by each agency which adapt the principles in way which is appropriate for their particular regulatory context.

Other forms of control mechanisms may also be used as well as rules and principles. These include strategies of oversight, of competition and of mutuality. These could take the form, for example, of performance assessment against a set of standards for the agency and individuals within it, league tables linked to agency funding, benchmarking projects, peer review, and the development of forums in which ideas of best practice can be exchanged. In developing standards for assessment, however, attention needs to be given to matching the performance criteria to the aims sought to be achieved (not using measurements based principally on sanctions if strategies of persuasion, education and conciliation are also seen to be important), and avoiding the 'league table syndrome': skewing activities to those parts of the job which are measured to the detriment of those which are not.

Principles of enforcement also need to be internalised, and this cannot be accomplished simply through the imposition of external controls. Attention needs to be given to ensuring that the

informal cultures within the organisation support the formal policies and processes. Principles, codes of practice or performance standards will not achieve the desired effect if compliance with them is simply a formal process and does not affect substantive behaviour. Internalisation involves a commitment to the principles, codes of practice and so on, and also knowledge of what they require (drawing on the shared understandings noted above), knowledge of how to implement them, and the organisational capacity to do so. That in turn may require, for example, education and the development of forums for the exchange of best practice ideas. Finally, the process of internalisation requires those practices to be made 'real': to be thoroughly integrated into organisational policy and culture. Embedding practices in this way requires frequent internal assessment of performance through self audit, and external assessment through audit verification and other forms of monitoring.

Finally it has to be remembered that none of these measures will alone or jointly guarantee success; they may however go some way to minimising the risk of failure.

Bibliography

- I. Ayres and J. Braithwaite (1992), *Responsive Regulation* (Oxford: OUP)
- E. Baar (1992), 'Partnerships in the Development and Implementation of Canadian Air Quality Regulation' 14 (1) *Law and Policy* 1-43
- E. Bardach and R. Kagan (1984), *Going By the Book: The Problem of Regulatory Unreasonableness* (Philadelphia, Penn.: Temple Press)
- R. Baldwin (1990), 'Why Rules Don't Work' 53 *Modern Law Review* 321
- R. Baldwin (1995), *Rules and Government* (Oxford: OUP)
- R. Baldwin and K. Hawkins (1984), 'Discretionary Justice: Davis Reconsidered' Public Law 580
- R. Baldwin and J. Houghton (1986) 'Circular Arguments: The Status and Legitimacy of Administrative Rules' Public Law 239
- D. Black (1976), *The Behaviour of Law* (New York: Academic Press)
- J. Black (1997a), *Rules and Regulators* (Oxford: OUP)
- J. Black (1997b), 'Which Arrow: Rule Type and Regulatory Policy' Public Law 94
- J. Black (1998a), 'Talking About Regulation' Public Law 77
- J. Black (1998b), 'Reviewing Regulatory Rules: Responding to Hybridisation' in J. Black, P. Muchlinski and P. Walker (eds), *Commercial Regulation and Judicial Review* (Oxford: Hart Publishing)
- J. Black (1999), 'Using Rules Effectively' in C. McCrudden (ed) *Regulation and Deregulation: Policy and Practice in the Utilities and Financial Services Industries* (Oxford: OUP)
- J. Black (2001 forthcoming), 'Dencentring Regulation: Understanding the Role of Regulation and Self Regulation in a "Post-Regulatory" World' in M. Freeman (ed), *Current Legal Problems 2001* (Oxford: OUP)
- J. Braithwaite (1985), *To Punish or Persuade: Enforcement of Coal Mine Safety* (Albany: New York Press)
- J. Braithwaite, P. Grabovsky and J. Walker (1987), 'An Enforcement Taxonomy of Regulatory Agencies' 9 *Law and Policy* 323
- Cabinet Office, Regulatory Impact Unit, 1999, 'Proposals to Improve the Quality of Enforcement Services: A Consultative Document on Section 5 of the Deregulation and Contracting Out Act 1994' (London: Cabinet Office)
- Cabinet Office, Regulatory Impact Unit (2000), *Good Policy Making: A Guide to Regulatory Impact Analyses* (London: Cabinet Office)
- Cabinet Office, Service First (1999), *How to Improve Local Authority Regulatory Services* (London: Cabinet Office, Service First, Effective Performance Division)
- Cabinet Office, Service First (2000), *Public Sector Excellence Programme* (London: Cabinet Office)
- CAB 197/99, 'Consultation Launched on Business Friendly Enforcement of Regulations' (London: Cabinet Office Press Release)
- W.G. Carson (1970), 'White Collar Crime and the Enforcement of Factory Legislation' 6 *Brit J Criminol* 383
- W.G. Carson (1970), 'Some Sociological Aspects of Strict Liability and the Enforcement of Factory Legislation' 33 *Modern Law Review* 396
- W.G. Carson (1982), *The Other Price of Britain's Oil* (Oxford: Martin Robertson)
- R. Cheit (1987), *Setting Safety Standards: Regulation in the Public and Private Sectors* (Berkeley: Univ Calif Press)
- M. Clarke (1990), *Business Crime: Its Nature and Control* (Cambridge: Polity Press)
- D. Cook (1989), *Rich Laws Poor Law* (Milton Keynes: Open University Press)
- R. Cranston, *Regulating Business* (London: Macmillan, 1979)
- K.C. Davis (1969), *Discretionary Justice* (Baton Rouge: Louisiana State University Press)

- J.F DiMento (1986), *Environmental Law and American Business: Dilemmas of Compliance* (New York: Plenum Press)
- C.S. Diver (1980), 'A Theory of Regulatory Enforcement' 28 *Public Policy* 257
- C.S. Diver (1983), 'The Optimal Precision of Administrative Rules' 93 *Yale Law Journal* 65
- R. Dworkin (1977), *Taking Rights Seriously* (Cambridge, Mass: Duckworth Press)
- Equality Commission of Northern Ireland (2000), *Guide to Statutory Duties: A Guide to the Implication of Statutory Duties on Public Authorities arising from s.75 of Northern Ireland Act 1998* (Belfast)
- L. Edelman, S. Petterson, E. Chambliss, S. Howard (1991), 'Legal Ambiguity and the Politics of Compliance: Affirmative Action Officers' Dilemma' 13(1) *Law and Policy* 73
- P. Fenn and C. Veljanovski (1988), 'A Positive Theory of Regulatory Enforcement' 98 *Econ J* 1055
- Financial Services Authority (2000) *A New Regulator for a New Millenium* (London: FSA)
- B. Fisse and J. Braithwaite (1993), *Corporations, Crime and Accountability* (Cambridge: Cambridge University Press)
- M. Friedland (ed) (1990), *Securing Compliance: Seven Case Studies* (Toronto: University of Toronto Press)
- J. Froud, R. Boden, A. Ogus and P. Stubbs (1998), *Controlling the Regulators* (Basingstoke: Macmillan)
- D.J. Galligan (1986), *Discretionary Powers* (Oxford: OUP)
- H. Genn (1993), 'Business Responses to the Regulation of Health and Safety in England' 15 *Law and Policy* 219
- W. Gray and T. Scholz (1991), 'Analysing the Equity and Efficiency of OSHA Enforcement' 13 *Law and Pol* 185
- P. Grabosky and J. Braithwaite (1986), *Of Manners Gentle* (Melbourne: Oxford University Press)
- N. Gunningham (1987), 'Negotiated Non-Compliance: A Case Study of Regulatory Failure' 9(1) *Law and Policy* 69
- N. Gunningham (1991), 'Private Ordering, Self Regulation and Futures Markets: A Comparative Study of Informal Social Control' 13(4) *Law and Policy* 297
- N. Gunningham and P. Grabovsky (1998), *Smart Regulation* (Oxford: OUP)
- F. Haines (1997), *Corporate Regulation: Beyond 'Punish or Persuade'* (Oxford: OUP)
- S. Halliday (2000), 'The Influence of Judicial Review on Bureaucratic Decision-Making' *Public Law* 110
- K. Hawkins (1984), *Environment and Enforcement* (Oxford: OUP, 1984)
- K. Hawkins (1989), 'FATCATS' and Prosecution in a Regulatory Agency: A Footnote on the Social Construction of Risk' 11(3) *Law and Policy* 370
- K. Hawkins (1990), 'Compliance Strategy, Prosecution Policy and Aunt Sally' 30 *British Jnl of Criminology* 444
- K. Hawkins (1992a), *The Uses of Discretion* (Oxford: OUP)
- K. Hawkins (1992b), 'The Use of Legal Discretion: Perspectives from Law and Social Science' in K. Hawkins (1992a), *The Uses of Discretion* (Oxford: OUP)
- K. Hawkins and B. Hutter (1993), 'The Response of Business to Social Regulation in England and Wales: An Enforcement Perspective' 15(3) *Law and Policy* 199
- K. Hawkins and J. Thomas (eds) (1984a), *Enforcing Regulation* (Boston, Mass: Kluwer-Nijhoff)
- K. Hawkins and J. Thomas (1984b), 'The Enforcement Process in Regulatory Bureaucracies' in K. Hawkins and J. Thomas (eds) (1984a), *Enforcing Regulation* (Boston, Mass: Kluwer-Nijhoff)
- C. Hood (1996), 'Control over Bureaucracy: Cultural Theory and Institutional Variety' 15(3) *Journal of Public Policy* 207
- C. Hood (1998), *The Art of the State* (Oxford: Clarendon Press)
- C. Hood, R. Baldwin and H. Rothstein (2000), 'Assessing the Dangerous Dogs Act: When Does Regulatory Law Fail?' *Public Law* 282

- C. Hood, C. Scott, O. James, G. Jones and T. Travers (2000), *Regulation Inside Government* (Oxford: OUP)
- B. Hutter (1988), *The Reasonable Arm of the Law?* (Oxford: OUP)
- B. Hutter (1997), *Compliance: Regulation and Environment* (Oxford: OUP)
- B. Hutter and S. Lloyd-Bostock (1990), 'The Power of Accidents: The Social and Psychological Impact of Accidents and the Enforcement of Safety Regulations' 12(2) *Law and Policy* 103
- B. Hutter and P. Manning (1990), 'The Contexts of Regulation: The Impact Upon Health and Safety Inspectorates in Britain' 12(2) *Law and Policy* 103-136
- B. Hutter and P. Sorenson (1993), 'Business Adaptation to Legal Regulation' 15(3) *Law and Policy* 169
- R. Kagan (1978), *Regulatory Justice: Implementing a Wage Price Freeze* (New York: Russell Sage Foundation)
- R. Kagan (1989), 'Understanding Regulatory Enforcement' 11(2) *Law and Policy* 89
- R. Kagan and J. Scholz (1984), 'The "Criminology of the Corporation" and Regulatory Enforcement Strategies' in K. Hawkins and J.M. Thomas (eds) (1984), *Enforcing Regulation* (Boston, Mass: Kluwer-Nijhoff)
- L. Lessig (1999), *Code and Other Laws of Cyberspace* (New York: Basic Books)
- M. Levi (1987), *Regulating Fraud: White Collar Crime and the Criminal Process* (London: Tavistock Publications)
- I. Loveland (1995) *Housing Homeless Persons* (Oxford: Oxford University Press)
- T. Lowi (1969), *The End of Liberalism* (New York: W.W. Norton, 1969)
- D. McBarnet and C. Whelan (1991), 'The Elusive Spirit of the Law' 54 *Modern Law Review* 848
- D. McBarnet and C. Whelan (1999), *Creative Accounting and the Cross Eyed Javelin Thrower* (Chichester: John Wiley and Sons)
- C. McCrudden (1999), 'Mainstreaming Equality in the Governance of Northern Ireland' 22(4) *Fordham Jnl of International Law* 1697
- C. McCrudden (2001), *Equality in Colin J. Harvey, Human Rights, Equality and Democratic Renewal* (Oxford: Hart Publishing)
- T. Makkai and J. Braitwaite (1993), 'Praise, Pride and Corporate Compliance' (1993) 21 *International Jnl of the Sociology of Law* 73
- T. Makkai and J. Braithwaite (1994a), 'Reintegrative Shaming and Regulatory Compliance' 32 *Criminology* 361
- T. Makkai and J. Braithwaite (1994b), 'The Dialectics of Corporate Deterrence' 31 *Jnl of Research in Crime and Delinquency* 347
- K. Manning (1988) 'Review of J. Braitwaite, *To Punish or Persuade*' 28(4) *British Jnl of Criminology* 559
- E. Meindinger (1986), 'Regulatory Culture: A Theoretical Outline' 9 *Law and Policy* 355-386
- P. Olsen (1992), *Six Cultures of Regulation* (Copenhagen: Handelshojsskolen)
- Office for Public Management (2000), *The Effectiveness of Different Mechanisms for Spreading Best Practice* (London: Cabinet Office)
- R. Nobles (2000), 'Enforcing Employee's Pension Rights - The Court's Hostility to the Pensions Ombudsman' *Industrial Law Journal*
- C. Parker (1997), 'Converting the Lawyers: The Dynamics of Competition and Accountability Reform' 33 *The Australian and New Zealand Jnl of Sociology* 39
- C. Parker (1999a), 'Compliance Professionalism and Regulatory Community: The Australian Trade Practices Regime' 26(2) *Journal of Law and Society* 215
- C. Parker (1999b), 'How to Win Hearts and Minds: Corporate Compliance Policies for Sexual Harrassment' 21(2) *Law and Policy* 21
- C. Parker (forthcoming), *The Open Corporation: Self Regulation and Corporate Citizenship*
- F. Pearce and S. Tombs (1990), 'Ideology, Hegemony and Empiricism' 30(4) *British Jnl of Criminology* 423

- M. Power (1997), *The Audit Society* (Oxford: Oxford University Press)
- J. Rees (1988), *Reforming the Workplace: A Study of Self Regulation in Occupational Safety* (Philadelphia: University of Pennsylvania Press)
- J. Rees (1997), 'Development of Communitarian Regulation in the Chemical Industry' 19(4) *Law and Policy* 477
- A. Reiss (1984), 'Selecting Strategies of Social Control over Organisational Life' in K. Hawkins and J.M. Thomas (eds) (1984) *Enforcing Regulation* (Boston, Mass: Kluwer-Nijhoff)
- R. Rhodes (1997), *Understanding Governance: Policy Networks, Governance, Reflexivity and Accountability* (Buckingham: Open University Press)
- G.M Richardson, A. Ogus and P. Burrows (1983), *Policing Pollution: A Study of Regulation and Enforcement* (Oxford: Clarendon Press)
- G.M. Richardson and M. Sunkin (1996), 'Judicial Review: Questions of Impact' Public Law 79
- P. Rock (1995), 'Sociology and the Stereotype of the Police' 22(1) *Jnl of Law and Society* 17
- N. Rose (1999), *Powers of Freedom* (Cambridge: Cambridge University Press)
- J. Rowan-Robinson, P. Watchman and C. Barker (1990), *Crime and Regulation: A Study of the Enforcement of Regulatory Codes* (Edinburgh: T&T Clarke)
- M. Samuels (1998) *Towards Best Practice: An Evaluation of the first two years of the public sector benchmarking project 1996-1998* (Cabinet Office, Next Steps Team, Efficiency and Effectiveness Unit)
- J. Scholz (1997), 'Enforcement policy and corporate misconduct: The changing perspective of deterrence theory' 60(3) *Law and Contemporary Problems* 253-268
- J. Scholz and W. Gray (1990), 'OSHA enforcement and workplace injuries: A behavioural approach to risk assessment' 3 *Jnl Risk and Uncertainty* 283-303
- W.A. Scott (1994), *Institutions and Organizations* (Thousand Oaks, Calif: Sage Publications)
- P. Selznick (1992), *The Moral Commonwealth* (Berkely: University of California Press)
- M. Shapiro (1988), *Who Guards the Guardians? Judicial Control of Administrative Action* (Athens: Univ Georgia Press)
- S. Shapiro (1984), *Wayward Capitalists* (New Haven, Conn: Yale University Press)
- D. Weil (1996), 'If OSHA is so bad, why is compliance so good?' (1996) 27 *RAND Jnl of Economics* 618
- D. Vogel (1984), *National Styles of Regulation: Environmental Policy in Great Britain and the United States* (Ithaca: Cornell University Press)
- P. Yeager (1991), *The Limits of the Law: The Public Regulation of Private Pollution* (Cambridge: Cambridge University Press)