OXFORD REVIEW OF ECONOMIC POLICY, VOL. 14, NO. 4

MAKING THINGS STICK: ENFORCEMENT AND COMPLIANCE

ANTHONY G. HEYES
University of London[1]

Environmental rules and regulations are only useful insofar as firms can be persuaded to comply with them—in full or in part. We survey the rapidly growing literature on the enforcement aspects of environmental policy. The difficulties facing any regulatory agency are likely to be exacerbated by information problems, penalty constraints, and the evasion efforts of firms, and we emphasize the role that recent innovative approaches to implementation can play in bolstering more traditional enforcement instruments.

I. INTRODUCTION

Environmental regulations are only useful if firms comply with them. Since compliance is generally costly, regulations have to be enforced if they are to work. Since enforcement is itself costly, enforcement is usually incomplete, meaning that some firms get away with non-compliance. Only by taking account of this sort of leakage can the success of a regulatory programme be assessed accurately.

While enforcement is an important dimension of any regulatory programme, the problems involved in enforcing environmental regulations are particularly pronounced. Ignoring enforcement issues in policy design and assessment here is likely to lead to particularly misleading results.

Non-compliance with many environmental regulations is commonplace and *effective* regulatory standards diverge substantially from the *nominal* standards of published legislation and agency directives. By its nature, evidence on non-compliance is scant, but some authoritative studies do exist. The US Government Accounting Office (GAO), for example, estimated that 65 per cent of regulated sources may be in violation of air pollution emissions limits (reported in Russell, 1990, p. 255). In the UK, the Environment Agency estimates that, depending upon region, compliance rates of trade dischargers with

[1] I am grateful to Tim Leunig, Chris Riley, Stephen Smith, and Dieter Helm for very helpful advice.

numerical discharge limits is 48–83 per cent, with an average of 74 per cent (see Department of the Environment, 1998). In most countries published rates of compliance with key environmental standards is less than full, often substantially less.[2]

In section II we present the standard economic theory of compliance and enforcement. The usual approach is to treat 'cheating' as a gamble such that the standard theory of choice under uncertainty can be applied. In the following sections we outline some of the ways in which the basic model can be extended to make it a more useful tool for policy design. These include taking account of self-reporting, criminalization of non-compliance, and the scope for the involvement of private citizens in enforcement.

The focus throughout is on general principles. While we draw on the USA and EU for examples to illustrate general points, it is not our aim to give detailed account of the institutions or procedures of enforcement in any particular country. Most of the principles presented apply across instruments, and through most of the discussion we take the choice of instrument (standards, permits, etc.) as given. We also ignore the fact that sensible policy-makers will calibrate their instruments in anticipation that there may be 'leakage' at the implementation—focusing on the enforcement process *per se*.

II. UNDERSTANDING COMPLIANCE

The first step to saying anything useful about enforcement strategy is to understand the way in which firms respond to enforcement incentives.

The model we will set up here is necessarily a very stylized one and it is worth noting that the relevance of any particular model of enforcement will depend crucially on the particular type of regulation involved. There is likely to be a world of difference between the types of programmes needed to ensure compliance with process, emissions, and environmental quality standards. Fly-tipping and other local misdemeanours in the waste field are very different from breaches of IPC authorizations, for example.[3]

How do firms decide whether or not to comply?

It is conventional for economists to model the compliance decision of a firm as a choice under risk—a gamble—with monitoring essentially a random process.

Suppose there exists some regulation requiring a firm to execute action a (e.g. to install a particular piece of abatement equipment, to stop emissions of a particular substance from a particular discharge pipe). If the cost to the ith firm of complying with that regulation is c_i, the probability of non-compliance being detected is π and the penalty for non-compliance is P, then it is apparent that—in the absence of other consideration—a profit-maximizing and risk neutral firm will comply if and only if

$$c_i \leq \pi.P. \qquad (1)$$

The right-hand side is the expected penalty for non-compliance, the left-hand side the firm's cost of compliance. Only those firms that find compliance sufficiently cheap will comply—the rest will take the risk of being caught and fined.

What matters in environmental terms is the compliance rate across all firms. It is plausible to think that in most settings firms will differ in how costly they find it to comply. This might reflect differences in

[2] True compliance rates—which are what really matters—are likely to be even less impressive. When official data say that 70 per cent of firms are compliant, what that really means is that for 70 per cent of firms the inspection agency has not established non-compliance. Given the inadequacy of most inspection programmes this is, obviously, a much less compelling statement. In one well-known study of the US Environmental Protection Agency (EPA) by the US GAO, conducted in 1979, it was found that of those sources which the EPA had designated as compliant with air emissions standards *only 3 per cent were actually compliant.* The distinction between actual and published non-compliance also gives rise to the paradoxical likelihood that as the intensity of any inspection programme is *reduced*, published rates of compliance will go *up*. A general principle in this field is: do not take published compliance figures too seriously.

[3] When operators must in practice install the equipment required to meet process standards, the non-compliance issue relates more to accidents—for example, resulting from inadequate quality control or process maintenance. The simple model developed here relates most closely to deterministic polluting technologies—though it can straightforwardly be generalized. There is an interesting related issue about how to design a regulatory system, with appropriate monitoring and enforcement regimes, that will ensure average discharges are acceptable while permitting periodic peaks. Stochastic pollution process throws up a number of tricky policy-design issues of this sort.

skills, corporate structures, plant locations, or technologies. If c is distributed according to some cumulative distribution $F(c_i)$, then the compliance rate across the industry as a whole (which we will call γ) can be expressed as a function of the enforcement policy parameters[4]

$$\gamma = F(\pi.P). \qquad (2)$$

It is obvious from looking at the inequality in equation (1) that raising the probability that non-compliance will be penalized and/or raising the size of the penalty will make compliance more attractive to the firm and so increase γ. The size of that increase—how effective raising π or P would be—will depend mainly upon the 'shape' of F.

What of the cost–benefit efficiency of the induced pattern of compliance? Assuming social welfare to the unweighted sum of industry costs and environmental damage, compliance decisions will be first-best if and only if the product $\pi.P$ happens to equal the marginal expected environmental damage caused by non-compliance. Such an expected penalty serves to internalize the externality due to the non-compliance and can be referred to as a 'Pigovian penalty'. For any *given* population compliance rate it is worth noting that the *distribution* of compliance effort between firms is efficient—it is always those firms with the lowest compliance costs that comply.

In a model of this sort the agency maximizes compliance (i.e. minimizes environmental damage) by setting both π and P as high as possible. It will only be able to ensure full compliance, however, if it can set them such that $\pi.P$ exceeds the upper bound of c. In most cases this will not be possible—budgetary, legislative, and other constraints almost invariably put limits on how high expected penalties can be raised.

While this sort of set-up makes modelling relatively easy, it is not particularly realistic. It assumes that the compliance decision faced by each firm is a binary one—a firm chooses either to comply or to violate, there is no 'halfway' option. Most real compliance decisions are, in fact, 'continuous' in character. This is true not just in the environmental setting but in many others. A motorist does not just choose between the options of 'speeding' and 'not speeding' but, rather, chooses exactly how fast to drive. A firm does not just choose whether or not to violate—it will typically have to choose a level of pollution which is an inherently continuous variable.

Suppose, to capture this, that firm i is subject to a regulatory standard which forbids it from discharging effluent e_i beyond some level S. Assume that the expected penalty for exceeding the standard is some increasing function $P(e_i - S)$ of the size of the violation,[5] and costs are increasing in environmental cleanliness according to a function $c(e_i)$. Then the firm's problem is to choose a level of emissions to minimize

$$c(e_i) + P(e_i - S). \qquad (3)$$

The solution to this problem (which we will call e_i^*) is implicitly defined by the associated first-order condition:

$$c'(e_i^*) = -P'(e_i^* - S). \qquad (4)$$

The firm pollutes up to the point at which the marginal cost of further abatement equals the marginal saving in terms of expected penalties.

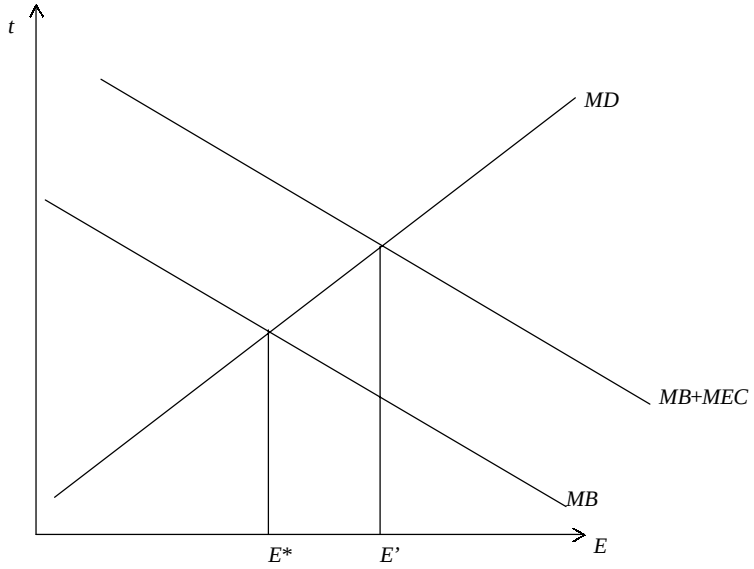
It is an important point to note that to solve equation (4) we do not have to know anything about the *level* of penalties (i.e. the value of $P(e_i^* - S)$), only about their properties at the margin. Once the decision to violate has been taken, the size of the violation depends only on the *marginal*, not the *average* properties of the expected penalty function—it is not the size of penalties that matters, but rather the 'speed' with which they increase with the degree of violation. This is the 'theory of marginal deterrence' (Shavell, 1992).[6]

[4] A cumulative distribution measures cumulative probability mass. In other words $F(x)$ tells us the proportion of firms for which $c_i < x$. It is apparent that $F(.)$ must lie between zero and one and must be everywhere (weakly) increasing in its argument.

[5] If the enforcement programme is prone to type I errors (falsely determining compliant firms to be non-compliant) then $P(e_i - S)$ may be positive even if $e_i < S$, i.e. even if the firm is complying (see Segerson, 1988).

[6] This is rather like the firm's choice of output level in industrial economics. To find that level we go to the point at which the marginal cost and marginal revenue functions cross. Fixed costs (and by implication total costs) do not have an impact on this decision because they do not affect anything *at the margin*.

Figure 1



Average and marginal penalties do not always move together—one enforcement regime may involve harsher penalties but have a 'flatter' penalty structure—and this can throw up some paradoxical results.

(i) The Optimal 'Amount' of Enforcement

Compliance models provide a link between changes in policy instruments and environmental performance. Equation (2), for example, describes a causal relationship from penalties and inspection probabilities to population compliance rates.

While very little empirical work has been done in the UK, a number of attempts have been made to operationalize models of this sort in the USA. Gray and Deily (1996), for example, use data on individual steel mills to study the relationship between EPA enforcement of air pollution regulations and firms' compliance decisions, while Cohen (1986) estimates the impact of US Coast Guard patrols on the frequency and severity of oil spills in US waters. Similar work has been done by Magat and Viscusi (1990) on pulp mills' compliance with water pollution regulations, by Feinstein (1989) on enforcement of safety regulations at US commercial nuclear power plants, and by Epple and Visscher (1984) on marine pollution.

The aim of each of these studies is to answer a question of the form: 'By how much will violations fall if enforcement parameter X is increased by 1 per cent?' Enforcement activity is (in most cases) costly, however, so cost-efficient policy design will require the balancing of the environmental benefits from enforcement (of the sort that such empirical studies characterize) against its costs.

In Figure 1, E on the bottom axis is the total discharge of some pollutant by an industry. The marginal damage (MD) and marginal industry benefit (MB) functions are drawn in the usual way, giving an optimal pollution level of E^* at the point where they intersect:[7]

$$MD(E^*) = MB(E^*). \qquad (5)$$

But achieving reductions in E implies not just costs to industry but also enforcement costs, EC. Adding marginal enforcement costs (MEC) to the diagram means that the optimal level of pollution—that which the enforcement agency should target—is E', the level which ensures that

$$MD(E') = MB(E') + MEC(E'). \qquad (6)$$

It makes sense that E' is greater than E^*—taking account of the enforcement costs associated with

[7] The MB slopes down under the assumption of diminishing marginal returns to abatement effort—the 100th unit of emission is less costly (in terms of spending on technology, or forgone production) to get rid of than is the 90th, and so on. Most classes of pollutant impose increasing marginal damage.

implementing a given outcome leads the policy-maker to 'water down' his or her objectives. In terms of the diagram, those increments of pollution between E^* and E' are socially inefficient according to standard cost–benefit criteria, but the administrative cost (in terms of a greater enforcement effort) of implementing a policy to remove them is excessive.

Taking account of enforcement costs, then, implies a 'distortion' in policy objectives away from that suggested by conventional analysis which tends to abstract from transactions costs. In some cases the regulatory agency may be prevented from optimizing its enforcement programme in this way:

> Sometimes the search for the optimum in enforcement priorities is blocked by statutory dictates: all establishments of a certain kind, the law might say, shall be inspected once every so often. Such directives may reflect legislative responses to catastrophes or scandals or may simply reflect notions of equal treatment. Either way, they can be very inefficient. (Bardach and Kagan, 1982, p. 165)

How great any such distortion might be will depend upon how sizeable those enforcement costs are in a given context, but it may be substantial. It is much more difficult to estimate *marginal*—as opposed to average—costs in this sort of setting, and this makes empirical optimization of enforcement programmes notoriously complex.[8,9]

A classic empirical analysis of the optimization of an enforcement programme is Cohen's (1986) study of the prevention of oil pollution in US coastal waters (Figure 1 here corresponds with Figure 1 in Cohen's paper). He uses detection-adjusted regression techniques to estimate the marginal costs of oil-spill prevention. The key results are summarized in Table X (p. 185) of that paper. The estimated *marginal* cost of preventing the release of one gallon of oil is estimated (all in 1986 dollars) to be \$5.50. Of this, however, only \$3.98 is the cost to the industry associated with additional preventive efforts (i.e. $MB(E) = 3.98$). The remaining \$1.52 (or 27.6 per cent of the total) is the enforcement cost associated with the US Coast Guard having to operate more frequent patrols (i.e. $MEC(E) = 1.52$). Clearly any attempt to set an optimal pollution target using cost–benefit criteria but ignoring *MEC*—given that it is so substantial—is likely to be well off the mark.

In the case of 'Superfund' in the USA—a programme designed to collect money from polluters of land to fund clean-up—a study by the Washington-based think tank, Resources for the Future, conjectured that for every dollar collected from polluters and used to clean up damaged land, as much as 50 cents could go on enforcement and transactions costs (principally the cost of tracking down those responsible and extracting the appropriate money from them). Even more alarmingly, Porter and van der Linde (1995, p. 115) quote a study that estimates that '88 per cent of the money spent by insurers between 1986 and 1989 went to pay for legal and administrative costs, while only 12 per cent was used for actual site clean-ups'.

It is also important to note that enforcement considerations could and should have an impact upon the choice of regulatory instrument—with the costliness of enforcement being anticipated at the point at which an instrument is selected, not just at the point at which a given instrument is calibrated.[10]

III. BEYOND THE BASICS

The type of model that we have developed here describes compliance behaviour—with its implication for enforcement strategy—in the simplest pos-

[8] In most areas of economics average things are easier to estimate than are their marginal counterparts. Think of a firm's costs of production, for instance. Average cost is simply total costs divided by total output, both of which can be observed directly. Marginal cost, on the other hand, is not based on observables and involves coming up with an answer to the hypothetical question 'by how much would my total costs have gone up had I chosen to produce one more unit of output?'

[9] Sometimes the costs or benefits need to be adjusted to account for particular industry circumstances—Gray and Deily (1991), for example, look at the considerations particular to the environmental regulation of a declining industry.

[10] It is instructive to note that the combination of a particular enforcement regime with some regulatory instrument A may come to yield a hybrid instrument which has the incentive properties more akin to some other regulatory instrument B. Thus a (non-binding) emissions limit s combined with an expected penalty p per unit of violation comes to mimic—in terms of its incentive properties—a linear emissions tax regime with $t = p$.

sible regulatory settings. The analysis is essentially borrowed from the more general economic analysis of law initiated as long ago as 1968 by Becker.

While the results of such work provide a useful bench-mark, the analysis needs to be extended. The simple model has a lot of unrealistic features and some of the more restrictive assumptions need to be relaxed if it is to be of much use in policy development. Some are purely technical—though none the less important to resolve—such as taking account of risk aversion in modelling the behaviour of both firms and regulators, taking account of the fact that firms may depart from naïve profit-maximizing behaviour, and so on. Some are more substantive.

(i) Self-reporting

Most environmental enforcement programmes do not simply involve the regulatory agency conducting random inspections of would-be polluters but, rather, include an element of self-reporting: 'Self-reporting is becoming an increasingly common feature of enforcement, particularly enforcement of pollution standards' (Livernois and McKenna, 1997, p. 1).

In this case understanding incentives can be quite a bit more complicated. The firm has to decide not just how much to emit, but also how much of any emissions to report (there can be no presumption that the firm will necessarily report honestly). The agency will no longer conduct inspections randomly but can, rather, condition those inspections on the content of the paperwork it gets from each firm, in very much the same way as the Inland Revenue can target its income tax audits on the basis of individual tax returns.

A number of authors have shown that such self-reports can greatly help effective enforcement. Harford (1987), Livernois and McKenna (1997) and others have shown that a number of the 'conventional wisdoms' generated by standard models do not necessarily carry over once account is taken of the strategic role of self-reporting. (Livernois and McKenna, for example, show that raising pollution penalties can increase pollution rates by interfering with the incentives for honest reporting.)

Of course, the reality of enforcement models with self-reporting is more complex than stylized models of this sort might suggest. A company's incentives are mixed—the raw data they hold are auditable, as is (in most cases) their compliance data process. Such failure might not only affect their public image, but would also expose them to additional penalties.

(ii) Noisy Monitoring

The 'gamble' model of the type sketched in section II is often motivated by the joint assumption of random inspection plus accurate inspection technology. In reality, monitoring equipment is likely to yield only a noisy estimate of actual emissions from a particular source such that the enforcement process is likely to be characterized by both type I and type II errors (i.e. with some actual violators being miscategorized as compliers and vice versa).

Recognizing the possibility of type I errors—that some firms may be deemed to be violators even when they are really compliant—leads to the possibility of 'more than full' compliance in settings where the regulated firm faces a continuous decision problem. That is, once firms have cut emissions down to the permitted level they may choose to abate even further to reduce the chance of being mistakenly prosecuted. This can be a source of 'regulatory chill' whereby regulations have greater than anticipated impacts and may even dissuade firms from operating in some sectors.[11] The risk of type I errors—the false prosecution of the innocent—is one of the most substantial arguments against the popular view that it would be desirable to force everyone to comply by setting penalties arbitrarily high (along with being one of the key reasons why legislators and courts are unwilling to allow regulatory agencies to impose draconian penalties).

Russell (1990) and others have argued that improving the accuracy of monitoring technology should be a key priority of environment agencies, and engineering research in this area has received increased agency funding in recent years.

One interesting possibility is that in many contexts there may be things that polluters can do to make

[11] Of course a smart regulatory agency might anticipate this and take account of it in calibrating the regulation.

themselves 'difficult' to inspect, making 'inspectability' a strategic variable for the firm.

In the USA a firm's constitutional right to privacy (under the 4th amendment) means that inspectors are obliged to conduct at least the initial rounds of their work from outside the firm's perimeter fence using remote-sensing devices. If the accuracy of such equipment decreases with distance, the firm can invest in uninspectability simply by buying more land—putting greater distance between the source of the pollutant and the nearest point from which detection can legally be attempted (see Strock, 1990). Alternatively, a firm might establish 'sanitized areas'—operationally redundant 'dummies' established for the benefit of inspectors. If the inspection process is seen as a sampling game in which the inspector tries to find a non-compliant part of the plant (the illegally set effluent outlet among the 20 properly set ones, for instance) then the firm can decrease the likelihood that he or she does so simply by increasing the number of sanitized areas. Linder and McBride (1984, p. 339) provide evidence of this and other sorts of 'attempt to change operations or employ idle capacity in order to pass on-site inspections'.

Heyes (1993) provides a formal model in which firms can avoid expected penalty by investing in 'uninspectability' as an alternative to spending on pollution control. A key result of that analysis is that because increasing the frequency of inspections encourages firms to switch towards less easily inspectable choices of technique, agencies should conduct *less frequent* but *more detailed* inspections than suggested by existing studies.

In a similar spirit Kambhu (1989) constructs a model in which a non-compliant firm can—by spending money on high-powered lawyers or in other ways—erode the penalty paid for a violation of given magnitude. In his model actual environmental performance can, intriguingly, be *negatively* related to the stringency of the regulatory requirement in place.

The uninspectability problem is probably less in the UK—where the Environment Agency has substantial rights of access and inspection—than it is in the USA.

(iii) Penalty Structures

While the simple, continuous framework described in equations (5) and (6) is rather general, it belies the difficulty in specifying the penalty function adequately.

In many cases the penalty function $P(.)$ can reasonably be treated as exogenous—not at the discretion of the enforcement agency. In others (such as is the case with the EPA in the USA) the agency exercises quite substantial discretion over the penalties levied and it is useful to think about what the optimal specification of $P(.)$ would be.[12] In the USA the penalty levied is explicitly broken into a component chosen to recover any economic gains from violation and a 'gravity' component, designed to be purely punitive.

EPAs in most countries, including the USA and Britain, are constrained—in some settings quite tightly—by legislation or the judiciary in how heavily they can fine non-compliant firms. It is not necessarily the case, however, that compliance-maximization will involve penalizing all violators to the fullest extent possible. More subtlety is required in specifying a profile of penalties which has good 'marginal deterrence' properties. The penalty faced by a firm following a minor infraction must not be so great that the firm has no incentive to prevent any release into the environment from escalating. Shaffer (1990) and others have done extensive and interesting work in this area.

In one recent paper, Heyes (1996) adapts the basic deterrence model to take account of the fact that pollutants differ in the extent to which their impacts are 'persistent'. Firms choose how much effort to exert in preventing an accidental spill, but, if a spill occurs, have also to choose whether or not to admit

[12] In the UK the regulatory agencies bring prosecutions, but fines are imposed by the Courts. As such, the UK Environment Agency has no scope to vary penalties—and hence to devise penalty structures with good marginal deterrence properties—though it may canvas the Courts. The possibility of introducing administrative fines—at the discretion of the Agency—could usefully be looked at further in the UK.

that it has happened and instigate clean-up. The instigation of clean-up is especially important if the pollutant is highly persistent (such as oil in the ocean, or radioactive pollutants with long half-lives).[13] The penalty regime needs to be harsh enough to ensure adequate incentive for prevention, but at the same time 'forgiving' enough to ensure that the firm responsible party does not skulk away once an accident has actually occurred. An interesting feature of Heyes's model is that the optimal penalty turns out to be *non-monotonic* in the persistence of the pollutant being regulated—with the highest penalties being reserved for firms discharging pollutants of 'medium' persistence.[14]

(iv) Judgement Proofness

In setting penalties it is not just higher powers that may prevent penalties being set as high as EPAs might like. It has often been noted by agencies and observers that a firm's asset base constitutes an upper bound on the penalty that can be levied—bankruptcy offers an escape hatch which means that a firm cannot be fined for more than its net worth. Shavell (1986) coined the term 'judgement proof' to describe such firms. This is likely to be a particularly worrying problem in two contexts: (i) where the regulated industry is populated by small firms (or, more accurately, by firms with comparatively small net worth) and (ii) in sectors where the type of environmental damage done is infrequent but catastrophic.

If the *effective* maximum penalty is restricted by the bankruptcy constraint, then the incentive properties of any particular enforcement regime can be compromised.[15] Where this problem is particularly pronounced the Agency may wish to compel firms to carry liability insurance.

The extent of judgement proofness, however, is likely to be anything but fixed—with firms recognizing the benefits of 'being small' when operating in particular sectors. The burgeoning rate of environmental penalty and litigation in the 1980s and 1990s has led to what Ringleb and Wiggins (1992) refer to as 'strategic subsidiarization', whereby large firms hive off environmentally sensitive components of their overall operations into independently incorporated subsidiaries with small asset bases, thereby protecting the assets of the parent company from exposure to environmental risk.[16]

(v) Multiple Polluters

Most early analysis of environmental enforcement—following its antecedents in law and economics—has assumed a single polluter.

In fact, in many settings several polluters will share an outlet. Many firms may, for example, discharge effluent into a particular stretch of river, or polluters may emit into the air in the same vicinity. This makes the task of inferring the contribution of a particular source from ambient measures difficult, in some cases impossible. In these cases the enforcement agency may be obliged to adopt 'second best' instruments and to regulate and monitor other signals of environmental performance, such as the consumption of polluting inputs by source (see Segerson (1988) for some discussion of these issues).

Even when not linked by pollutant-recipient, firms may be strategically linked through the enforcement process itself—e.g. if the agency is budget-constrained and only has the resources to pursue the worst 'tail' of polluters.

[13] An example of a completely *unpersistent* pollutant would be something like noise which has only an impact effect and offers no meaningful opportunity for clean-up. The nature of biological pollutants is particularly interesting—the scope for breeding may mean that, rather than diminishing, the size of their impact actually grows through time.

[14] A *monotonic* relationship means that as one thing goes up so does the other. Non-monotonic in this context means that as we increase the persistence of the pollutant in question then, other things being equal, the optimal penalty at first rises then falls.

[15] This is not that unusual—very many firms in high-tech sectors, for examples, may have the potential to do environmental damage of much greater value than their monetary net worth. Small biotech firms handling biological pollutants or haulage firms engaged in the transportation of hazardous waste provide examples.

[16] Ringleb and Wiggins (1992) provide compelling evidence of the occurrence of strategic subsidiarity in the USA. There remains some doubt over the efficacy of the strategy. Courts may choose to 'pierce the corporate veil'—penalize the parent for the actions of the subsidiary—if the subsidiarization is judged to been a ploy purely designed to avoid legal sanction. There is, as yet, insufficient case history to predict to what extent—and in what sorts of cases—courts will be willing to do this, and how far those decisions will be sustainable under appeal.

(vi) Criminalization

One of the most high-profile developments in environmental enforcement in the USA since the mid-1980s—one not followed in the EU—has been the increased use of *criminal* sanctions.

The type of analysis presented in section II above has effectively treated penalties for environmental infractions as being pecuniary and incident upon the firm. The firm is assumed to treat such penalties as a 'cost of doing business' and to treat their minimization in very much the same way as it would treat the minimization of any other cost.

Criminalization provides another dimension to the EPA's armoury—the penalty for wrongdoing is no longer an entry in the corporate accounts. The Agency, in conjunction with the Department of Justice, can pursue individual employees within firms—those with responsibility for environmental management—with a view to holding them criminally liable for environmental damage. Every year in the USA many dozens of executives are tried and imprisoned for the environmental damages of the firm that employs them. The EPA in the USA puts great emphasis on the availability of this as a weapon. Though the numbers of convictions remain comparatively small they may be disproportionately important—the recognition of individuals that *they themselves* may face prison for the failings of the firm for which they work having a substantial impact. As Russell (1990, p. 264) notes:

> A third line of enforcement effort is a new stress on criminal prosecution. Since penalties for conviction then include time in prison, and since the effort has been accompanied by a strong PR exercise, it appears that the EPA is striving to plant fear in the hearts of executives deciding whether or not to comply with environmental regulations.

The extent to which identifying and attributing blame to particular individuals within organizations is likely to be effective in generating incentives will depend upon a variety of factors, including the extent to which there are likely to exist specific individuals within a firm who have the controls necessary and sufficient to determine environmental performance (these have been investigated by, among others, Tietenberg and Segerson, 1994). In the longer term one might expect a continued trend towards criminalization to encourage corporations to restructure job descriptions to ensure that no individual (say of board level) has 'responsibility' for environmental compliance. While this may serve to avoid the risk to personnel of prosecution, the pursuant muddying of lines of environmental control may not be socially desirable.

Criminalization appears to have lost favour as a 'way ahead' in the UK and elsewhere in the EU in recent years.[17] The role that it could and should play remains, however, one of the most debated.

(vii) Privatization of the Enforcement Function

It is inevitable that private individuals will have an impact on compliance incentives—they may influence Agency policies through political channels, and their responses are likely to underlie the so-called 'market incentives' whereby firms behave well to avoid losing customers, employees, and investors. This said, however, it is usually taken for granted that government—through one or more of its agencies—will have monopoly over regulatory enforcement.[18]

The most striking and potentially far-reaching shift in emphasis in the USA in the 1990s, however, has been the increasing frequency with which individual citizens and citizen groups have intervened in the enforcement process directly. This is a trend which has not been mirrored in the UK or the rest of the EU. Before 1970, the state and federal agencies held exclusive enforcement responsibility in the USA. In 1970, Congress amended the Clean Air

[17] Much (almost all) of UK environmental regulation is underpinned by criminal law. The difference in style between the USA and UK is in the application of criminal penalties. In the UK very minor offences—which in the USA would be treated as administrative offences—can be criminalized, but there is comparatively little use of major criminal penalties (imprisonment) for bigger offences. What may be needed in the UK is not an increase in criminalization but a more selective approach, focusing on more substantial breaches.

[18] Note that we are not talking here about situations where individuals sue firms to gain compensation for damage inflicted against them personally—a type of action familiar in tort law—but rather where individuals or private groups move to enforce a piece of regulation more generally.

Act to allow private parties to pursue non-compliant firms which the EPA fails to pursue. Since then similar provisions have been built into other statutes, including the Clean Water Act and Toxic Substances Control Act (ToSCA). The number of citizen suits has increased dramatically since the early 1980s and now represents a substantial proportion of the total.[19] Citizen involvement in the process in the UK is much more limited—despite persistent pressure from some of those groups for channels to be widened. Individuals and groups can in theory take out a prosecution against polluters in the UK, but will not in general be able to force agencies to reveal data or information beyond what is on public registers. It is more likely that public groups such as non-governmental organizations (NGOs) will seek judicial review.

The desirability of direct private involvement in enforcement is open to question. As Naysnerski and Tietenberg (1992) note: 'While the role of NGOs in environmental policy is growing rapidly, our analytical understanding of the consequences of this emerging role has not kept pace.' They put forward what is probably the most straightforward view of the impact of private involvement: 'Adding the likelihood of private enforcement action to that of public enforcement implies a higher probability that a non-compliant firm will be penalized which . . . should increase the observed degree of compliance with the regulation' (Naysnerski and Tietenberg, 1992, p. 43).

According to such a view public and private enforcement efforts are additive. There are several reasons, however, for thinking such a view may be too simplistic. While it holds in a world in which the public agency operates a random but incomplete enforcement programme, as soon as one allows for the possibility that the EPA may do anything more subtle than this it breaks down. When the agency is exploiting penalty leverage (as in Harrington, 1988) or engages in regulatory dealing (as in Heyes and Rickman, 1998)—both of which will be explained a little later—the decision not to pursue a particular violation is a strategic decision by the EPA and may be compliance-enhancing such that intervention by a private enforcer could be expected, in general, to compromise the overall efficacy of the programme. Similarly—and depending upon the budgetary process by which the agency's budget is set—private enforcement could lead to a cut in funds for the public enforcement effort and so have a deleterious effect overall.

None of which is to say that private enforcement effort is necessarily a bad thing—just that it needs more consideration. The opposition by agencies in some countries to the extension of private enforcement rights should not necessarily be interpreted cynically.[20] In addition, insofar as direct private involvement in the enforcement of environmental regulations is to be privatized, more attention should be paid to how 'efficient' levels and patterns of such activity can be encouraged through fiscal or other means (see Heyes, 1997).

IV. WHY DO FIRMS COMPLY SO MUCH?

While the aim in this paper is not to survey the large body of empirical research which has developed in this field, it is worth noting a particular empirical regularity which has been identified in a variety of contexts by a variety of authors.

Winston Harrington (1988, p. 29) and others have noted—in the context of all of the major enforcement programmes operated by the EPA in the USA—that despite the fact that:

(*a*) when the EPA observes violations it often (almost always) chooses not to pursue the violator and

[19] In the early 1980s '[H]igh rates of non-compliance with the water pollution and other laws generated private enforcement at a level not before seen in American regulatory law' (Yaeger, 1991, p. 320). In the 5 years before 1983 private groups filed only 41 lawsuits under the water law, in 1983 the number was 103, with 87 in the first-quarter of 1984 alone. 'This activity, much of it organized by various national environmental groups, began to rival the federal government's own enforcement action: of the 108 actions in 1983, 62 eventuated in actual citizen lawsuits, compared to the 77 suits filed by the Department of Justice on behalf of the EPA' (Yaeger, 1991, p. 321). For some excellent legal discussion see Fardil (1985).

[20] Boyer and Meidinger (1985, p. 841) assert that '[t]he agencies resist private enforcement in the belief that the plaintiff groups are intruding on bureaucratic turf'.

(*b*) the expected penalty faced by a violator who is pursued is small compared to the cost of compliance, it is still the case that

(*c*) firms comply a significant proportion of the time.

While the exact terms differ by context, the underlying and recurrent pattern is the same: firms appear to over-comply—to comply more fully and/or more frequently than would be suggested by consideration of the private costs and benefits of so doing. The same pattern has been noted outside the USA—see Heyes and Rickman (1998), Hawkins (1983), and a variety of the citations listed.

This so-called Harrington paradox is perhaps the best known empirical 'result' in this field, and various commentators have provided alternative rationales for it—we list five. The apparent puzzle provides a good context within which to think about most of the important issues in enforcement.

So why might it be that firms would seem to comply too much?

Voluntary compliance

So far we have assumed that firms are cynical profit-maximizers. It is sometimes contended that there is in fact such a thing as a 'green corporation' which has a social conscience and attaches weight to its environmental performance *per se*. The main problem with such a theory is evolutionary—a firm that forgoes profit to pursue other objectives (green or otherwise) is likely to find itself displaced in the market by one that does not. Alternatively, in contexts where there are barriers to entry in the product market, the disciplinary function of the market for corporate control is to ensure that managers who fail to maximize shareholder value will come to be replaced by others who do.[21] Authors such as Arora and Cason (1996) have provided evidence on voluntary compliance (in that case with the EPA's so called 33/50 programme in the USA) but that 'voluntary' behaviour can be explained in term of other benefits to compliance which generally go unmeasured.

Misjudgement

It may be that polluters overestimate the probability that wrong-doing will be detected or the penalties that such detection would trigger. There is compelling evidence in the public finance literature that many private individuals misjudge the probability that their tax return will be audited—maybe a similar mechanism is at work here. There is, however, no good survey evidence to support such a conjecture. There is anecdotal support for the notion that individual company employees may overestimate the probability of criminal prosecution (a probability which, despite the growing number of highly publicized cases, remains minuscule).

Unmeasured costs to violation

Numerical analyses of compliance behaviour routinely assume that the cost of being found non-compliant is simply the administrative or litigative penalty which must be paid. There may, however, be additional 'market' penalties which firms face. The profitability of firms may be adversely affected through the responses of customers, investors, and employees to a poor environmental record. Badrinath and Bolster (1996) estimated that 86 per cent of the penalty for environmental prosecution in the USA is reputational (a higher fraction in the case of violations of the Clean Air Act). Grabosky (1994) and others have emphasized the loss in brand image among consumers and the loss of morale among employees that prosecution may imply.[22] While policies to make information about the environmental performance of firms more widely available—so that consumers, investors, and employees can bring pressure to bear on the 'bad'—will likely work in the right direction of making firms more likely to behave socially, they should not be seen as a panacea. Even in a world in which everyone is fully informed about everything, the fundamental externality problem—that when I consume a product the manufacture of which has caused environmental damage, most if not all of that damage is incident upon others—remains.

'Penalty leverage'

The basic model of compliance around which this paper has been built has been static. Harrington

[21] The evidence is that the emergence of the green or ethical investor in recent years had had little impact on the overall market for corporate control.

[22] As Friemann (1995, p. 362) notes: 'Taking care of the social and ecological consequences of corporate activities, even if they promise no immediate financial gains, may turn out to be an element of a modern far-sighted management strategy for a variety of reasons.'

(1988) and others have emphasized the repeated nature of the interaction between firm and agency. Once repetition is taken into account, the agency is likely to condition its attitude towards a particular firm on that firm's past performance—much as a criminal court will take account of an individual's previous behaviour in determining a sentence. Harrington shows that in such a model optimal behaviour will exhibit apparent over-compliance at any *given* moment (or in any *given* repetition of the game). The argument is refined in Harford (1991).[23]

'Regulatory dealing'

Heyes and Rickman's (1998) model of regulatory dealing is also consistent with the Harrington Paradox. The model takes account of the fact that EPAs typically interact with a given firm in more than one context (the firm might have several plants, operate in several different geographical locations, or be subject to several different sets of environmental regulations). In this case, there is scope for the Agency to exploit 'issue-linkage'. In such a world, firms may *appear* to over-comply in a given setting, but in reality are so doing in exchange for the agency 'turning a blind eye' somewhere else (at another plant, or in its enforcement of some other regulation). This sort of story fits well with the case study and anecdotal evidence of people such a Keith Hawkins (in the UK) and Peter Yaeger (in the USA) who have spent time inside regulatory agencies and find that they engage in various sorts of 'horse trading' of this sort.

The explanations, of course, are neither exhaustive nor mutually exclusive. As usual in the real world the relative importance of different effects will depend upon a range of factors and upon context.

V. CONCLUSIONS

Regulations are only useful insofar as they are enforced—either fully or partially. Current rates of compliance with many of the most important pieces of environmental legislation in the USA, EU and elsewhere are so low that the *effective* stringency of regulation is determined as much by the intensity of enforcement efforts and expenditures as it is by the stringency of the legislative decree itself.

Enforcement—the implementation phase—is the nitty-gritty of environmental regulation and one on which policy analysts should and do put increasing emphasis. The aim here has been to provide an idea of the richness of issues involved. An important thing to bear in mind in thinking about these things is how different the various types of pollution are. Noise and radioactive emissions throw up, it is obvious, very different control problems and we should not be surprised if the correct response is a non-uniform one—with different enforcement solutions being best in different control situations.

This said, it is useful—with one eye on the current British context—to draw a few general policy conclusions.

- Enforcement costs and limits must be anticipated in both the choice of regulatory instrument and its calibration.

- Optimal targeting of inspection effort will require conditioning on both the characteristics of the source, and the source's compliance history. The form of such conditioning may be highly complex.

- Administrative control over penalties—allowing, in the British case, the Environment Agency to set penalties—is likely to be beneficial. The agency can have regard not just to the case in isolation but to the incentive properties of the penalty structure as a whole, in particular its marginal deterrence properties.

- Selective use of custodial criminal sanctions for violations at the 'upper end' may well have an important incentive role to play in deterring errant firms from treating financial penalties as a 'cost of doing business'.

[23] As well as conditioning monitoring strategies on the past compliance performance of firms, an agency can also condition those strategies on observable characteristics of firms in a non-random way—identifying the characteristics associated with non-compliance across a population and drawing up risk-profiles of offenders. This is similar to the type of approach routinely used by Customs and Excise, for example, in countering VAT fraud. The Environment Agency in the UK has a new Operator and Pollution Risk Assessment (OPRA) model to help it target inspection in this way.

- Agencies should be wary of private intrusion in the enforcement process. This is not to say that private prosecution of regulatory non-compliance may not have a role to play in some contexts—merely that its inherently uncoordinated nature may interact adversely with the incentives produced by a well-designed public enforcement programme.

The two most intriguing trends in the USA—ones which have been largely resisted in the UK—are towards criminalization of non-compliance, and the move towards the *de facto* 'privatization' of parts of the enforcement function. The extent to which these are precedents that should be followed—and if so in what settings—is up for debate.

REFERENCES

Arora, S., and Cason, T. N. (1996), 'Why do Firms Volunteer to Exceed Environmental Regulations? Understanding Participation in EPA's 33/50 Program', *Land Economics*, **72**(4), 413–32.

Badrinath, S. G., and Bolster, P. (1996), 'The Role of Market Forces in EPA Enforcement', *Journal of Regulatory Economics*, **10**(2), 165–81.

Bardach, E., and Kagan, R. A. (1982), *Going by the Book: The Problem of Regulatory Unreasonableness*, Philadelphia, Temple University Press.

Becker, G. (1968), 'Crime and Punishment: An Economic Approach', *Journal of Political Economy*, **76**, 169–217.

Boyer, B., and Meidinger, E. (1985), 'Privatising Regulatory Enforcement: A Preliminary Assessment of Citizen Suits Under Federal Environmental Laws', *Buffalo Law Review*, **34**(3), 833–940.

Cohen, M. A. (1986), 'The Costs and Benefits of Oil Spill Prevention and Enforcement', *Journal of Environmental Economics and Management*, **13**, 167–88.

Department of the Environment (1998), *Digest of Environmental Statistics 1998*, London, The Stationery Office.

Epple, D., and Visscher, M. (1984), 'Environmental Pollution: Modelling Occurrence, Detection and Deterrence', *Journal of Law and Economics*, **27**, 29–60.

Fardil, A. (1985), 'Citizen Suits Against Polluters—Picking up the Pace', *Harvard Environmental Law Review*, **9**(1), 23–84.

Feinstein, J. S. (1989), 'The Safety Regulation of US Nuclear Power Plants: Violations, Inspections and Abnormal Occurrences', *Journal of Political Economy*, **97**(1), 115–54.

Friemann, J. (1995), 'Environmental Information Systems and Eco-auditing', in H. Folmer, H. Landis Gabel, and H. Opschoor (eds), *Principles of Environmental and Resource Economics*, Cheltenham, Edward Elgar.

Grabosky, P. N. (1994), 'Green Markets: Environmental Regulation by the Private Sector', *Law and Policy*, **16**(4), 419–48.

Gray, W. B., and Deily, M. E. (1991), 'Enforcement of Regulation in a Declining Industry', *Journal of Environmental Economics and Management*, **21**, 260–24.

— — (1996), 'Compliance and Enforcement: Air Pollution Regulation in the US Steel Industry', *Journal of Environmental Economics and Management*, **31**, 96–111.

Harford, J. D. (1987), 'Self-reporting of Pollution and the Firm's Behaviour Under Imperfectly Enforceable Regulations', *Journal of Environmental Economics and Management*, **14**, 293–303.

— (1991), 'A Reconsideration of Enforcement Leverage when Penalties are Restricted', *Journal of Public Economics*, **45**, 391–5.

Harrington, W. (1988), 'Enforcement Leverage When Penalties are Restricted', *Journal of Public Economics*, **37**(1), 29–53.

Hawkins, K. (1983), 'Bargain and Bluff: Compliance Strategy and Deterrence in the Enforcement of Environmental Regulations', *Law and Policy Quarterly*, **5**(1), 35–73.

Heyes, A. G. (1993), 'A Model of Regulatory Enforcement when Inspectability is Endogenous', *Environmental and Resource Economics*, **4**, 479–94.

— (1996), 'Cutting Pollution Penalties to Protect the Environment', *Journal of Public Economics*, **60**(2), 251–65.

— (1997), 'Environmental Regulation by Private Contest', *Journal of Public Economics*, **61**(2), 407–28.

— Rickman, N. (1998), 'A Theory of Regulatory Dealing—Revisiting the Harrington Paradox', *Journal of Public Economics*, forthcoming.

Khambu, J. (1989), 'Regulatory Standards, Compliance and Enforcement', *Journal of Regulatory Economics*, **1**(2), 103–14.
Linder, S. H., and McBride, M. E. (1984), 'Enforcement Costs and Regulatory Reform: The Agency and Firm Response', *Journal of Environmental Economics and Management*, **11**, 327–46.
Livernois, J., and McKenna, C. (1997), 'Truth or Consequences: Enforcing Pollution Standards', Department of Economics, University of Guelph, Canada, mimeo.
Magat, W. A., and Viscusi, K. (1990), 'Effectiveness of the EPA's Regulatory Enforcement: The Case of Industrial Effluent Standards', *Journal of Law and Economics*, **33**, 331–60.
Naysnerski, W., and Tietenberg, T. (1992), 'Private Enforcement of Federal Environmental Law', *Land Economics*, **68**(1), 28–48.
Porter, M. E., and van der Linde, C. (1995), 'Towards a New Conception of the Environment–Competitiveness Relationship', *Journal of Economic Perspectives*, **9**(4), 97–118.
Ringleb, A. H., and Wiggins, S. N. (1992), 'Liability and Large-scale, Long-term Hazards', *Journal of Political Economy*, **98**, 574–95.
Russell, C. S. (1990), 'Monitoring and Enforcement', in P. Portney (ed.), *Public Policies for Environmental Protection*, Washington, DC, RFF.
Segerson, K. (1988), 'Uncertainty and Incentives for Nonpoint Pollution Control', *Journal of Environmental Economics and Management*, **15**, 87–98.
Shaffer, S. (1990), 'Regulatory Compliance with Nonlinear Penalties', *Journal of Regulatory Economics*, **2**, 99–103.
Shavell, S. (1986), 'The Judgement Proof Problem', *International Review of Law and Economics*, **6**(1), 45–58.
— (1992), 'A Note on Marginal Deterrence', *International Review of Law and Economics*, **12**, 133–49.
Strock, J. M. (1990), 'Protecting the Environment: EPA After 20 Years', *Trial*, August, 22–7.
Tietenberg, T., and Segerson, K. (1994), 'The Structure of Penalties in Environmental Enforcement: An Economic Analysis', *Journal of Environmental Economics and Management*, **23**(2), 179–200.
Yaeger, P. (1991), *The Limits of the Law: The Public Regulation of Private Pollution*, Cambridge, Cambridge University Press.