Voluntarism in the Shadow of Coercion

Anthony Heyes, University of London*

May 10th 1999

Abstract

Individuals and firms operating with the possibility of regulatory intervention hanging over them are often observed to behave in socially responsible ways. In the US, for example, many polluters have signed-up to environmental 'Voluntary Agreements' (VA's). In contrast to conventional wisdom, this paper shows that under realistic informational assumptions VA's will be made by firms with *high* compliance costs. This has a variety of unconventional implications. The analysis also points to the strategic role that VA's might play as explicitly *interim* measures. Whilst the environmental application is topical, the model can also be used to explain pro-social behaviour in settings as diverse as product markets, organisations and student common-rooms. **Keywords:** Environmental regulation - voluntarism

1 Introduction

The most striking innovation in environmental policy since tradeable permits has been the emergence, in the 1990's, of 'voluntary agreements' (VA's). VA's take various forms, but have in common that firms agree to verifiable improvements in environmental performance without coercion.[1]

*The author can be contacted at Department of Economics, Royal Holloway College, University of London, Egham Hill, Egham, Surrey TW20 0EX, England.

[1]What constitutes 'coercion' and 'voluntarism' is, as usual, ambiguous. Firms in our model are - as throughout the literature - self-serving. Reductions will be voluntary, but not altruistic.

The highest profile VA schemes have been designed and operated in the by the Environmental Protection Agency in the US, where they are now an established part of the regulatory landscape. The USEPA's 'Partners for the Environment' programme comprises 21 such schemes (detailed at http://www.epa.gov/partners).

What has been achieved under some of the voluntary programs is remarkable. The 33/50 program, for example, achieved its target of a 50% reduction in discharges of 17 toxic substances (versus a 1988 baseline) a year ahead of schedule.[2]

Why do firms agree to do something for (apparently) nothing? Various motives for participation in VA schemes have been suggested. In Segerson and Miceli (1998), firms and the regulatory agency arrive at a VA against the background threat that a regulation will be imposed if a voluntary arrangement is not secured. The assumption driving the result that a VA will always emerge in equilibrium is the assumption that legislation implies a transactions or compliance cost premium. A VA generates a surplus (the transactions and compliance costs saved by avoiding the legislative process) the division of which the EPA and firm bargain over.[3]

In some contexts there may be 'PR' returns to VA's through improved employee morale or customer morale, in others they may include cost-sharing incentives, with the EPA providing subsidy, either in cash or in kind. Frey (1997), in one of a series of papers, promotes 'intrinsic motivation' as driving voluntary pro-social behaviour in environmental and other settings. Significantly for policy, he presents evidence that such voluntarism can be crowded-out by coercion.

In this paper we concern ouselves with the role that voluntarism might play as a signalling device in contexts where an activity, though not regulated now, may be subject to regulation in the future. The EPA's voluntary program is a comparatively recent phenomenon, and the extent to which its

[2]Khanna and Damon (1999) warn against attributing the whole of the headline figure to the program itself: 40% of the total reduction took place in the two years between 1988 (the baseline) and 1990 (when 33/50 was initiated), and whilst the releases of participants fell by 41% between 1990 and 1993, those of non-participants fell by 18%. Despite these qualifications, the achievements of the program still look substantial.

[3]Significantly, in the absence of such a cost differential there would be no advantage (nor disadvantage) to VA's for either the firm or the agency. In their model the assumed outcome in the case of failure to agree is implementation of an (inefficient) command-and-control regime and this creates a threat point against which bargaining can occur.

parts will ultimately come to be supplanted by formal regulation remains to be seen. In some cases, however, it is clear. The 17 toxic substances covered by the high-profile 33/50 program are also amongst the 189 chemicals listed under Title III of the 1990 Clean Air Act Amendments, under the terms of which the EPA will impose quantitative emissions limits starting in 2001 (United States Environmental Protection Agency (1990)). The case in which VA's are used as an explicitly interim device (*i.e.* where future regulatory intervention is certain) is nested as a special case in the model here.

We show that firms which have high costs of abatement can use pro-social behaviour during the pre-regulatory phase to signal type. By so doing they secure correspondingly laxer treatment should regulation arise later. It is important to note that this is not a model in which the firm 'buys' goodwill from the regulatory agency - the regulatory agency always acts in a time-consistent manner and so cannot be expected to reciprocate such goodwill in future periods.[4]

The implications of the analysis are several-fold, and will be drawn out more fully below. The prediction that VA's will tend to attract *high* cost firms is an interesting one, and counter to existing models. Existing studies attempting to predict participation in VA's empirically do not allow the validity of this to be assessed. Extrapolating from data on realised abatement costs gathered from VA's to project what the compliance cost implications of industry-wide regulation might be will yield an *over*-estimate. This is in contrast to the existing conventional wisdom which is that VA's will tend to attract firms for whom participation is relatively *un*costly - those who find participation sufficiently cheap - and that such inference will therefore yield *under*-estimates.

The sorting role that VA's can play may also make them peculiarly suitable for use as short-term or interim devices in the lead up to more conventional regulation, and we investigate this in some detail in Section 3.

Whilst the application to environmental policy is particularly topical, the model could be used to explain pro-social behaviour in a plethora of market and non-market settings.

[4]The motivation for VA's is, as well, quite diffferent to that in Segerson and Miceli (1998), which is a complete information model.

2 A model

Consider a two period model populated by a large number of firms. In each period firms may engage in some anti-social activity (which we will refer to as emissions of a pollutant) which imposes external damage at a constant marginal rate b.

The cost of cutting back on that activity ('abatement') will be denoted $c(a|\theta)$ where $c' > 0$, $c'' > 0$ (there are diminishing returns to abatement effort). The parameter $\theta_i \in \{H, L\}$ is a firm-specific cost parameter, which is observed privately by the firm. The proportion of high cost firms in the population, γ, is common knowledge. High cost firm's have both higher absolute and marginal abatement costs (almost) everywhere: $c(a|H) > c(a|L)$, $c'(a|H) > c'(a|L)$, $\forall a > 0$. This implies that $(c(a|H) - c(a|L))$ is increasing in a - whilst low cost firms always enjoy a cost advantage, that advantage increases as the amount of abatement done increases.

In the first period the anti-social activity is not regulated. We follow Segerson and Miceli (1998) in assuming that at the beginning of the second period there is some exogenous probability π that a framework of regulation will be put in place. If so, that regulation will consist of a vector of optimally set (and fully-enforced) firm-specific emissions limits $(s_1, \ldots, s_i, \ldots, s_n)$. With probability $(1 - \pi)$ the activity remains unregulated in period 2.

That π is exogenous is a big assumption. Miceli and Segerson (1998: 112) provide political justifications - intervention might depend upon the election of a particular political party in a forthcoming vote, or upon space becoming available on the legislative slate (both of which might reasonably be treated as exogenous). In the model here the exogeneity is compatible with the small firm assumption - every *individual* firm treats the probability of emergence of a regulatory programme as insensitive to its own actions. Analytically, in setting-up the model we were keen to have π as exogenous in order to abstract from the incentives for industry to try to influence the likelihood of future regulation. That the ability to so-influence could provide another motive for voluntarism in an alternatively specified model is obvious. The case in which $\pi = 1$ - in which VA's are used an an explicitly interim measures - is nested as a special case in the curent model, and is considered in some detail in Section 3.

2.1 Optimal contingent regulation

In the event of regulatory intervention, the regulatory agency will select s_i (the standard imposed on the ith firm) to maximise expected net benefits:

$$b.s_i - E(c(s_i|\theta_i)). \tag{1}$$

(It is obvious from the structure of the problem that firm's will never choose to overcomply in period 2). In the absence of any information having been aquired in period 1, the subjective distribution of θ_i is simply the prior such that the expression in 1 becomes

$$\gamma.W(s|H) + (1-\gamma).W(s|L)$$

where we define the notation

$$W(s|\theta_i) \equiv b.s - c(s|\theta_i).$$

We will denote as s_γ - implicitly defined by

$$\frac{\gamma.c'(s_\gamma|H) + (1-\gamma).c'(s_\gamma|L))}{b} = 1 \tag{2}$$

- the value of s that maximises this, and assume throughout that $s_\gamma > 0$. (Convexity of costs implies that the associated second-order condition is satisfied). A uniform standard is set based on the 'average' firm, such that each firm abates upto the point at which the marginal benefit of abatement equals the (expected) marginal cost.

If the regulator knows a firms type, on the other hand, optimal regulation implies setting s to maximise $W(s|\theta_i)$. The first-order condition associated with an interior solution is:

$$W'(s|\theta_i) = 0 \Rightarrow \frac{c'(s|\theta_i)}{b} = 1. \tag{3}$$

We will denote by s_{θ_i} the implicit solution to Equation 3 conditional on a known θ_i. That $s_L > s_H$ - under full-information a low-cost firm faces a tougher abatement standard than does a low cost counterpart - follows directly from the assumed convexity of c and is unsurprising.

Note that the firm does not play an interesting role in period 2. If a regulation is imposed upon it, it complies.[5]

[5]Enforcement is assumed to be complete. Note that this implies that actual emissions are assumed verifiable by the regulatory agency. The firms compliance expenditures are assumed unverifiable. It is this latter assumption that prevents the agency implementing a single-period separating mechanism backed by penalties.

2.2 Voluntary Agreements?

Emissions are unregulated in period 1. The assumed exogeneity of π means that there is no scope to influence the *probability* of regulatory intervention. Furthermore, the assumed optimality of regulation (if it is introduced) means that there is no scope to attempt to 'capture' or otherwise influence the regulator through demonstrations of goodwill.[6]

Our contention is that a sub-set of firms can use pro-social behaviour in the lead up to possible regulation - voluntary abatement in period 1 - as a credible signal of type. Restricting attention to symmetric equilibria, we will denote by a_L and a_H the levels of abatement chosen by firms according to type. The regulatory agency's posterior beliefs are represented by the conditional probability distribution $\mu(\theta_i|a_i)$.

Of interest to us is here is whether we can generate an equilibrium other than $a_L = a_H = 0$ - one involving voluntary reductions in emissions. The equilibrium concept to be applied throughout is that of *Perfect Bayesian Equilibrium* (PBE). This involves specifying first period-abatement levels, the regulator's contingent response function and beliefs $\{a_{\theta_i}, s(a), \mu(\theta_i|a_i)\}$ such that (a) $s(a)$ is a best response mapping for the agency, given beliefs, (b) a_{θ_i} maximises the expected pay-off of a firm of type θ_i given the regulatory response mapping and (c) $\mu(\theta_i|a_i)$ is consistent with Bayes' rule.[7]

2.2.1 Signalling equilibria

Separation in this two type model implies $a_L \neq a_H$, $\mu(\theta_i|a_i) = 1$. The representative high and low cost firms choose different levels abatement and the regulator is able to infer a firm's type from its choice.

It is apparent from the structure of the problem that a low-cost firm

[6] With commitment one could imagine a contract (implicit or explicit) between regulator and firm in which the firm engages in voluntary abatement in period 1 'in exchange for' lenient treatment in period 2. We will call this the 'exchange of favours' story. It has been used by Hunt and Auster (1990), amongst others, to explain corporate environmentalism. Such *quid pro quo* cannot happen in the model presented here - the regulator is assumed sequentially rational and, moving second, would renege on any such agreement at the beginning of period 2.

[7] For a more formal definition of a PBE see Gibbons (1992). For the reader unfamiliar with PBE and related refinements (in particular the Intuitive Criterion) Tirole (1988: 436-450) provides an excellent presentation.

prefers non-separation, since $s_L > s_\gamma > s_H$ implies

$$c(s_L|\theta_i) > c(s_\gamma|\theta_i) > c(s_H|\theta_i), \forall \theta_i$$

A known L-type faces more stringent regulation than either a known H-type, or a firm about which the regulator has no information beyond his prior. Certainly the L-type will never invest in a costly signal of type (such as first period voluntary abatement).

If signalling is to occur, then, it must be by high-cost firms. This is an interesting realisation in its own right and a contrast to the usual mind-set of familiar signalling models (the Spence model and derivatives of it) in which it is the *more* able player attempting to distinguish herself.

Consider a candidate separating equilibrium in which a high-cost firm is able to distinguish itself by agreeing to a voluntary level of abatement a_V in period 1 (such that $a_H = a_V$), the low-cost firm doesn't mimic (such that $a_L = 0$), and posterior beliefs can be represented $\mu(H|a) = 1$ if $a = a_V$, $\mu(H|a) = 0$ otherwise.[8] Given such separation the regulator will (if regulation indeed goes ahead) impose an emissions ceiling s_H on any firm that subscribed to the VA in period 1, a tougher ceiling of s_L on any firm that didn't.

Establishing the existence of such a separating equilibrium requires identifying a value of a_V satisfying two incentive compatibility constraints. The representative high-cost firm has to want to participate in the VA. Its (best available) alternative is to do zero abatement in period 1 and be (wrongly) identified as a low cost firm, such that it will wish to participate if and only if $a_V < \widehat{a}$ where

$$c(\widehat{a}|H) = \pi.(c(s_L|H) - c(s_H|H)). \tag{4}$$

The left-hand side is the cost of the voluntary abatement. The right-hand side is the expected gain from successfully signalling type.

The no mimic condition - which ensures that the representative low-cost firm will not choose to participate in the voluntary reductions - requires that $a_V > \widetilde{a}$ where

$$c(\widetilde{a}|L) = \pi.(c(s_L|L) - c(s_H|L)). \tag{5}$$

[8]The presentation here is rather informal - and will be tightened up later in the paper. In particular we are being cavalier about out-of-equilibrium beliefs. Whilst what we are sketching is indeed a PBE it should be apparent that these beliefs imply that it will not satisfy the Intuitive Criterion.

This ensures that the gains to a low-cost firm from signing up to the VA (and hence sending a false signal) - the right-hand side - is outweighed by the costs (the left).

Thus:

Proposition 1 *If $\widehat{a} > \widetilde{a}$ then there exists a continuum of separating PBE's involving high (and only high) cost firms entering into VA's in the first period. These are described by*

$$a_H = a_V \ , \ \widehat{a} > a_V > \widetilde{a};$$

$$a_L = 0;$$

$$s(a) = s_H \ \textit{if} \ a > a_V, \ = s_L \ \textit{otherwise;}$$

$$\mu(H|a) = 1 \ \textit{if} \ \widehat{a} > a_H, \ = 0 \ \textit{otherwise.}$$

The only one of these that survives the elimination of weakly dominated strategies is $a_V = \widetilde{a}$.

Proof: Given the beliefs specified, a_H, a_L and $s(a)$ are obvious by construction. It is apparent that the posterior probabilities sum to unity, and are consistent with Bayes Rule. That elimination of weakly dominated strategies yields $a_V = \widetilde{a}$ is implied by the fact that $c(a|H)$ is everywhere increasing in a.

The signalling equilibria are best understood from Figure 1. First, $\widehat{a}$ is defined by the intersection of $c(\widehat{a}|H)$ and $\pi.(c(s_L|H) - c(s_H|H))$, the left and right-hand sides of Equation 4. The H-type firm will want to participate in the signalling equilibrium as long as a_V is not too great (provided, more concretely, it is less than $\widehat{a}$). Second, $\widetilde{a}$ is defined by the intersection of $c(\widetilde{a}|L)$ and $\pi.(c(s_L|L) - c(s_H|L))$ (which, note, lies strictly below $\pi.(c(s_L|H) - c(s_H|H))$, as drawn). The L-type will not attempt to mimic the signal provided $a_V > \widetilde{a}$, *i.e.* provided the obligations implied by so doing are sufficiently large (costly).

Thus, to profitably and credibly signal its type the high-cost firm must choose a voluntary level of abatement such that $\widehat{a} > a_V > \widetilde{a}$. If $\widehat{a}$ is strictly greater than $\widetilde{a}$ then there are a range of possible values of a_V, each generating a separating PBE. Eliminating weakly dominated strategies yields $a_V = \widetilde{a}$ - the least cost separating equilibrium (the H-types choose the cheapest signal that works). This his equilibrium satisifes the (somewhat stronger) Intuitive Criterion (Cho and Kreps (1987)).

Note that if $\widehat{a} < \widetilde{a}$ there are no such separating equilibria, and we assume that the outcome $a_H = a_L = 0$ will result.[9]

It is worthwhile noting the following:

Remark 1 *Insofar as signalling is a motive for VA's, the terms of those VA's - in particular the level of voluntary abatement a_v - cannot be understood from analysis of the cost circumstances of the VA participants.*

Where VA's exist, whilst H-types are the signatories, a_v equals $\widetilde{a}$ - implicitly defined by Equation 5 - and is wholly defined by the cost conditions faced by the *non-participating* firms. Once dominated strategies are eliminated the 'no mimic' condition is the binding constraint.

This has interesting implications for applied work (such as trying to understand motives for VA participation and design using case studies) in the area.

Remark 2 *Insofar as signalling is a motive for VA's, extrapolation of the cost of industry-wide abatement from the costs incurred by VA participants will lead to over-statement.*

This, again, has interesting implications for policy-analysis, and is the opposite of the conventional view that VA's will tend to attract firms with comparatively low costs (for whom participation is correspondingly uncostly) such that extrapolation would lead to *under*-statement.

3 VA's as a deliberately interim measure

The absence of regulation in period 1 has, thus far, been assumed. Regulation at the start of period 2 has, similarly, been assumed to emerge with some exogenous probability.

Let us assume, instead, that the EPA has regulatory discretion in each period. The Agency could, then, impose emissions levels in period 1.

Define W^γ as follows -

$$W^\gamma \equiv \gamma W(s_\gamma|H) + (1-\gamma)W(s_\gamma|L)$$

[9] As usual in models of this sort it is possible to construct other pooling equilibria characterised by $a_H = a_L > 0$. These are straight-forward to refine away, however, and do not interest us here. See Tirole (1988: 448) for standard analysis.

- the welfare generated by a period of 'untailored' regulation. Since $s_\gamma > 0$ it is true by construction that $W^\gamma > 0$. It must also be the case that $W(s_H|H)$ and $W(s_L|L)$ are positive. Thus, at the start of period 2 the EPA will always choose to regulate and will either set a uniform standard s_γ or tailored standards s_H and s_L on H- and L-type firms respectively.

The problem facing the EPA, then, is whether to regulate in period 1, or to postpone and give VA's a chance to work in period 1. Postponement can, then, be regarded as a strategic choice on the part of the Agency.

If setting emissions in both periods the Agency can do no better than set a uniform standard s_γ in both periods. Ignoring discounting, this yields two-period expected welfare equal to:

$$2W^\gamma \tag{6}$$

It if postpones regulation, relying on VA's in the interim, two-period expected welfare will be

$$\gamma W(\widetilde{a}|H) + (\gamma W(s_H|H) + (1-\gamma)W(s_L|L)) \tag{7}$$

if $\widehat{a} > \widetilde{a}$,

$$W^\gamma \tag{8}$$

otherwise. The latter is the case in which VA's do not emerge in equilibrium (recall Proposition 1) such that postponing regulation simply means that no abatement is done in period 1. Since $2W^\gamma > W^\gamma > 0$, when specification is such that $\widehat{a} \not> \widetilde{a}$ the EPA will regulate in both periods (there will be no role for VA's).

The interesting case is the former, in which postponement of regulation leads to the emergence of VA's in the interim. Combining 7 and 8 implies that it will be welfare-improving to postpone regulation if and only if

$$\gamma W(\widetilde{a}|H) + (\gamma W(s_H|H) + (1-\gamma)W(s_L|L)) > 2W^\gamma.$$

Rearrangement allows us to summarise in the following way:

Proposition 2 *A regulatory agency maximising undiscounted two-period welfare will rely on VA's in period 1 if and only if*

$$(\gamma W(s_H|H) + (1-\gamma)W(s_L|L)) - W^\gamma > W^\gamma - \gamma W(\widetilde{a}|H) \tag{9}$$

Both sides of the inequality are unambiguously positive, but their relative size is ambiguous. The right-hand side here is the period 1 cost of postponement. It is the difference between W^γ and the in-period benefit resulting from the VA's that would emerge under postponement. The left-hand side is the benefit of postponement - the gains deriving from being able to tailor second period regulation. The condition says, simply, that the Agency will postpone if the latter exceeds the former.

Note that any increase in W^γ - the welfare derived from a period of untailored regulation - makes postponement less likely. This is intuitive. It is apparent from the structure of the model that the inequality in 8 will not be satisfied for $\gamma = 0$ or $\gamma = 1$. The continuity in γ of both sides of the inequality leads directly to:

Proposition 3 *A regulatory agency maximising undiscounted two-period welfare will not postpone regulation if γ is sufficiently small or sufficiently large.*

This, again, is intuitive. For values of γ sufficiently close to either 0 or 1 there is little heterogeneity in costs across firms. Untailored regulation works comparatively well and postponing becomes less attractive, both because the short-run cost is higher and the long-run gains are lower (W^γ appears, correspondingly, on both sides of 9).

Insofar as signalling is a motive for VA's, then, voluntary programs will be more likely to feature in contexts where compliance costs are more heterogenous (in the sense of a sufficiently interior value of γ). The extent to which this yields a testable hypothesis will depend upon the extent to which firm-level cost heterogeneity can be measured.[10]

3.1 A digression on 'managed' VA's

In the analysis so far we have assumed that VA's emerged spontaneously - they are industry-initiated. In particular, the participating firm sets a_V.

Observation of the VA programme as it has developed in the US in the 1990's, however, is that the terms of VA's are often EPA-initiated. Under the 33/50 program, for example, the EPA designed a VA (requiring a 33% reduction in toxic emissions within 2 years, 50% within 5) and wrote to about 600 firms inviting them to participate. Khanna and Damon (1999: 11) note

[10] By assumption costs are not observeable *ex ante*. There is no particular reason to think that realised firm-level compliance costs will, in general, be observeable *ex post*.

that out of the 123 eligible firms in the chemical sector, for example, 44 signed-up immediately, 75 within 2 years.

In this case, then, it appears appropriate to treat the EPA as setting a_V, and offering it on a take-it-or-leave-it basis to polluters. We will refer to these as 'managed' VA's. For reasons to be spelt out below, we continue to regard industry-initiated VA's as the most worthy of analysis. It is straight-forward, however, to restate Proposition 2 for the case in which VA's are managed:

Proposition 4 *A regulatory agency maximising undiscounted two-period welfare will rely on managed VA's in period 1 if and only if*

$$(\gamma W(s_H|H) + (1-\gamma)W(s_L|L)) - W^\gamma > W^\gamma - \gamma W(a^{MANAGED}|H) \quad (10)$$

where $a^{MANAGED}$ *solves*

$$\max W(a^{MANAGED}|H)$$

$$\textit{subject to } \widehat{a} > a^{MANAGED} > \widetilde{a}.$$

If the VA in the case of postponement is firm-initiated it will, as we have noted, be characterised by voluntary abatement $a_V = \widetilde{a}$. If the H-type firm chooses to send any signal at all, it will set it the lowest value compatible with its being credible.

Under managed VA's of the sort envisaged here, however, the EPA would be able to set a_V to maximise the in-period welfare generated from participation, subject to the pair of incentive compatibility constraints required for separation.[11]

If the $a^{MANAGED}$ solving the constrained maximisation in the Proposition turns out to equal $\widetilde{a}$ then the ability to manage makes no difference, otherwise it makes the use of VA's strictly more attractive (the second stage benefits from being able to tailor regulation are unchanged - *i.e.* the left-hand sides of the inequalities in 9 and 10 are identical). Thus:

Remark 3 *The ability to manage VA's makes it (weakly) more likely that the EPA will postpone regulation. It will be strictly more likely if* $W'(\widetilde{a}|H) > 0$.

[11] That these are satisified must be the case. If the managed VA's are not going to induce separation - and hence allow better targetting of regulation in period 2 - then the regulator is better off not to use VA's at all, opting instead for two periods of untailored regulation and a 2-period pay-off $2W^\gamma$.

(Note that $W'(\widetilde{a}|H) > 0$ is a sufficient, though not necessary, condition for $a^{MANAGED}$ to strictly exceed $\widetilde{a}$).

We are, as already noted, more interested in understanding spontaneous than managed VA's, despite the apparent preponderance of the latter. This is because any $a^{MANAGED}$ stipulated in an EPA-designed VA document would typically be non-binding. Thus for any $a^{MANAGED} > \widetilde{a}$ an interested firm could agree to participate, so having its emissions monitored, but then under-achieve - *i.e.* go ahead and do $\widetilde{a}$ - still sending a credible signal. The exception to this would be if the VA was itself backed by an enforcement programme, such that having (voluntarily) agreed to participate a firm would then be required to honour the agreement.[12] We are unaware of any VA's which have an enforcement dimension of this sort - though there is nothing conceptual to rule it out.

4 Conclusions

Firms often find themselves having to operate in an environment where their activities may be subject to regulatory control in the future. In deciding how to behave today, the forward-looking manager should take account of how behaviour today could affect regulation tommorrow.

If firms have private information about compliance costs then *high*-cost firms can use voluntary pro-social behaviour in the pre-regulatory phase to credibly signal their type. The model predicts that, in equilibrium, firms that participate in voluntary pro-social behaviour can expect more lenient treatment later. This is a standard type of prediction, but is no longer dependent upon the 'exchange of favours' argument which does not stand up to the test of time consistency (the Agency should always renege on the deal).

The extent to which signalling might have provided one of the motivations - of which there must doubtless be many - for participation in existing programmes is difficult to judge. Regulatory experience remains short, empirical work in the field is scant, and the signalling decision is also likely to be sensitive to a variety of unobserveables - *e.g.* abatement costs, firm-level expectations regarding the probability of future regulatory intervention.

[12] This could be called the 'golf club' approach. No-one has to join, but if you do, you are then bound by the terms of membership (with compliance enforced by threat of forfeiture of a deposit or in other ways).

The model we present here, however, *is* consistent with the small empirical literature that has so far emerged.[13]

The model here suggests an alternative and novel motive for voluntarism and augments rather than replaces other explanations. It provides a new rationale for why somebody who is willing to 'do a bit' voluntarily might be exempted from having to 'do a lot' mandatorily (and, conversely, why those who refuse to do even a bit of their own volition may expect a tough mandate, should a mandate arrive). It might provide an alternative perspective on voluntarism in a variety of settings - environmental and other - where other explanations seem inapplicable.

5 Bibliography

Arora, S. and T. N. Cason, 1995. "An Experiment in Voluntary Environmental Regulation: Participation in EPA's 33/50 Program", *Journal of Environmental Economics & Management* 28: 271-286.

Bernheim, D. B. 1986. "On the Voluntary and Involuntary Provision of Public Goods", *American Economic Review* 76: 789-793.

Cho, I. K. and D. Kreps, 1987. "Signalling Games and Stable Equilibria", *Quarterly Journal of Economics* 102: 179-222.

European Commission, 1996. *On Environmental Agreements*, Communication from the Commission to the Council and European Parliament, Brussels.

Frey, B., 1997. "The Cost of Price Incentives: An Empirical Analysis of Motivation Crowding-Out", *American Economic Review* 87(4): 746-755.

Gibbons, R. 1992. *A Primer in Game Theory*, Harvester-Wheatsheaf: Hemel Hempstead, England.

Heyes, A. G., 1999a "Honesty in a Regulatory Context - Good Thing or Bad", forthcoming *European Economic Review*.

Heyes, A. G., 1999b. "Voluntarism in the Shadow of Coercion: An Application to Environmental VA's", mimeograph, Royal Holloway College, University of London.

[13]The best empirical evaluations so far have been Arora and Cason (1995) and Khanna and Damon (1999)). Neither of these is framed in such a way as to shed much light on our key testable hypotheses - though something at least tangentially applicable can be drawn out of each of them with appropriate reinterpreation of variables (this is discussed in more detail in Heyes (1999b)).

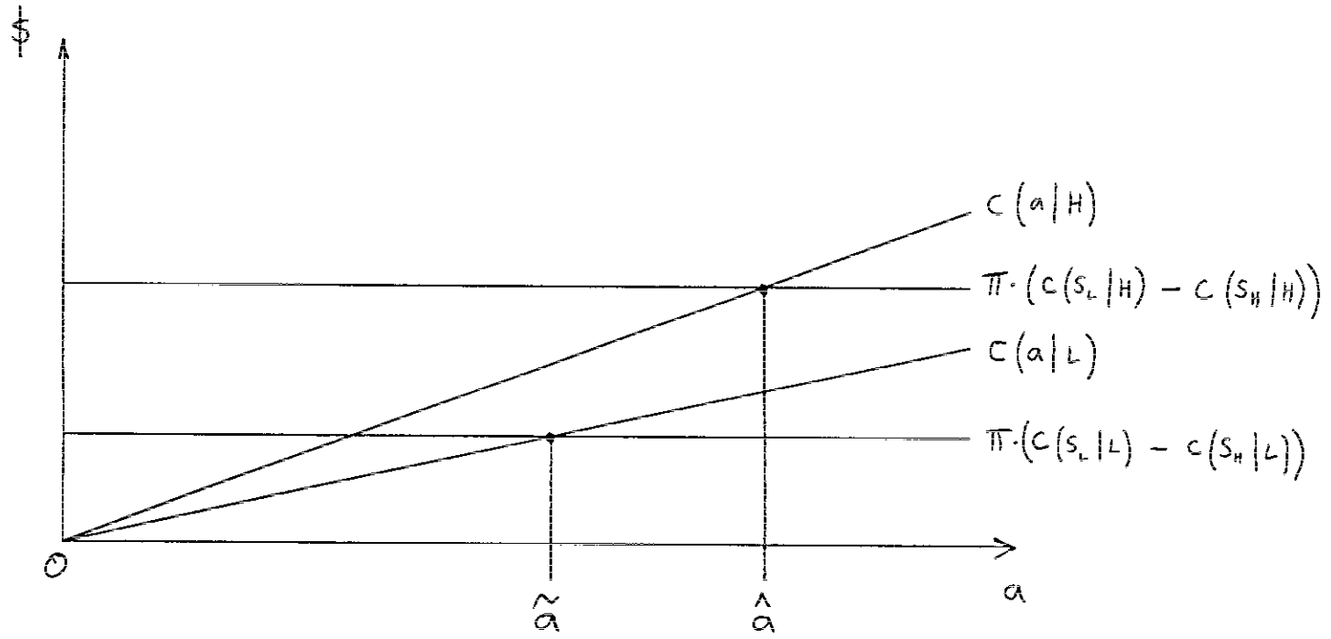
Hunt, C. B. and E. R. Auster, 1990. "Proactive Environmental Management: Avoiding the Toxic Trap", *Sloan Management Review*, 31: 7-18.

Khanna, M. and L. A. Damon, 1999. "EPA's Voluntary 33/50 Program: Impact on Toxic Releases and Economic Performance", *Journal of Environmental Economics & Management* 37: 1-25.

Segerson, K. and T. J. Miceli, 1998. "Voluntary Environmental Agreements: Good News or Bad for Environmental Protection", *Journal of Environmental Economics & Management* 36: 109-130.

United States Environmental Protection Agency, 1990. *The Clean Air Act Amendments of 1990 Summary Material*, Office of Air and Regulation, Washington DC.

Tirole, J. 1988. *A Theory of Industrial Organisation*, MIT Press: Cambridge MA.


$
c(a|H)
π·(c(s_L|H) − c(s_H|H))
c(a|L)
π·(c(s_L|L) − c(s_H|L))
0
ã
â
a