Community Pressure for Green Behavior*

Anthony Heyes†
University of Ottawa

Sandeep Kapur‡
Birkbeck, University of London

April 2012

Abstract

The desire to avoid rousing community hostility may encourage firms to behave in an environmentally responsible manner. Firms may engage in corporate social responsibility (CSR) to maintain community support and/or to regain the support of a community where it has been lost. It has been conjectured that such 'informal regulation' could effectively replace formal intervention in some settings, and usefully complement it in others. We explore these conjectures with mixed results. Informal regulation is necessarily less efficient than a well-designed formal alternative and the pattern of green behavior induced by the threat of community hostility may increase or decrease welfare. The existence of community pressure may increase or decrease the optimal calibration of a formal intervention (in this case an environmental tax) and may complement or detract from the incentives generated by an optimally-calibrated tax.

Keywords: community pressure, informal regulation, corporate social responsibility

*We are grateful to Simon Skene, Ron Smith, seminar participants at University College London, two referees and an editor of this journal for helpful comments and advice.

†Heyes (corresponding author) is at the Department of Economics, University of Ottawa, 55 Laurier Avenue East, Ottawa, K1N 6N5, Canada. *email*: aheyes@uottawa.ca, *tel*: +1-613-562-5753, *fax*: +1-613-562-5999.

‡Department of Economics, Birkbeck, University of London, Malet St, London, WC1E 7HX, United Kingdom. *email*: s.kapur@bbk.ac.uk, *tel*: +44-207-631-6405.

1 Introduction

In many settings the formal incentives for a firm not to damage the environment are weak, either because legal strictures are not in place or are ineffectually enforced. There may still, however, be substantial incentives for firms to exercise restraint in order to avoid community hostility:

> When formal regulation is weak or absent, communities can often use other channels to force pollution abatement by local factories in a process of "informal regulation" (Pargal and Wheeler (1996: 1314))

Much recent work has pointed to the scope of informal regulation, particularly in developing economies where formal governance structures are typically weak. van Rooij (2010a) investigates the extent to which community pressure can serve as an alternative to traditional command and control methods as a mechanism for pollution regulation in lower- and middle-income countries. But similar informal incentives are equally apparent in richer economies where the rationale for recent disclosure programmes in the United States and Europe, for example, is that they may improve environmental compliance. Various commentators have suggested that such disclosure programmes could be a viable alternative to formal regulatory intervention in some settings, or could complement such interventions (see, for example, Tietenberg (1998)). We explore both hypotheses here.

Pargal and Wheeler (1996) provide an early and important contribution to the economic analysis of informal regulation. They develop a market-type model for environmental services in a community. The polluting firm relies on supply of services from the community (employees, contractors, etc.) and/or on demand for its services from that community. The terms on which these inputs and outputs are traded are sensitive to community hostility. A firm that is known to damage the local environment may face a hostile community and so find it harder or more expensive to attract and motivate workers, for example, or harder to sell its products, generating informal 'penalties' for poor environmental performance.

The focus of recent case study and empirical work has been on estimating the strength of informal incentives and understanding the mechanics of such community pressure –

in other words the linkages from firm behavior to community hostility, and onward from community hostility to informal penalties. "Without recourse to legal enforcement of existing regulations (if any), they must rely on the leverage provided by social pressure on workers and managers, adverse publicity, the threat (or use) of violence against the plant, and pressure through politicians, local administrators, or religious leaders" (Pargal and Wheeler (1996: 1315)). 'Community' can mean different things in different contexts, and is not necessarily defined by geography. Customers (actual or prospective) may, for example, be widely dispersed, as may investors. These are both groups that designers of disclosure programs have sought to co-opt as potential 'informal regulators'.

Informal regulation is likely to have most impact if community hostility translates into reduced profits, through increased cost, decreased revenue, or both. In that case profit-motivated firms will choose socially-desirable actions in order to maintain community support. The label 'corporate social responsibility' (CSR) has been used to mean a variety of things in recent years, but the instrumentalist or strategic view of CSR is that profit-maximizing firms engage in CSR as an investment in reputational capital.[1] Lyon and Maxwell (2008) and Reinhardt et al. (2008) provide some useful discussion and context on the interpretations of CSR. The notion of doing things to maintain a supportive community is similar to the idea of stakeholder management in the business literature (Kassinis and Vafeas (2006)).

Dasgupta et al. (2000) examine data on plant-level compliance with air quality standards in Mexico and conclude that 'extra legal' factors (their term for community pressure) play an important role in compliance decisions, sometimes even leading to *over*-compliance. Wang (2000) estimates compliance incentives in a Chinese context and concludes that "... the implicit price from community pressure is at least as high as the explicit price (in the form of the pollution levy)". Hettige et al. (1996) provide empirical support for the impact of informal regulation in Asia, noting that "... despite weak or

[1]The firms in our model will be profit maximizers rather than altruists. Baron (2001) is amongst those who contend that the CSR label should only be applied when the motivation for good deeds is altruistic, not part of a profit-maximizing strategy ("The perspective taken here is that both motivation and performance are required for actions to receive the CSR label" (Baron (2001:9)). CSR in our paper will be defined by performance – all firms in our model maximize expected profit streams. This is the more usual use of the term amongst economists.

non-existent formal regulation, there are many clean plants in South and Southeast Asia". Evidence from developed economies includes Foulon et al. (2002) in Canada and Pargal et al. (1997) for the United States.

Environmental performance may be linked to informal penalties through a variety of channels. An environmentally dirty firm may be penalized in the labor market (the terms on which it can hire workers – e.g. Brekke and Nyborg (2008)), capital market (the terms on which it can raise capital – e.g. Badrinath and Bolster (1996)) and the product market (reduced demand for its outputs – Roe et al. (2001)). The product market premium for superior environmental performance has received particular attention recently and there are extensive literatures on 'green marketing' (see Becker-Olson et al. (2006) for a survey), green premia and eco-labels (Pelsmacker et al. (2006), Kotchen (2006)). Furthermore, poor environmental performance can lead to local or wide-spread consumer boycotts (Klein et al. (2004), Innes (2006)). It may also induce 'direct action' such as vandalism (van Rooij (2010b)).

Our objective in this paper is to consider the efficiency of the incentives that the threat of community hostility generates. The environmental performance of a firm and community attitude towards that firm can be expected to interact through time. The behavior of the firm may depend upon community attitude, whilst community attitude will itself be sensitive to that firm's choice of actions. A community might turn against a firm it observes as being environmentally irresponsible or 'dirty', but the firm may be able to repair its reputation by subsequent clean behavior (see, for example, Parsons (2011)). Since community attitude and firm behavior change through time it is natural to develop a dynamic model to explore their evolution.

We present a model in which at any given moment community attitude towards a particular firm is either 'hostile' or 'supportive'. We formulate a Markov-type regime-switching model in which this community attitude evolves through time in a way that is sensitive – probabilistically – to the firm's environmental choices. In the basic version of our model we assume that community attitude evolves purely in response to the firm's actions in the most recent period. This is a common modeling simplification in Markov models. In an extension we show that the qualitative insights of the model are sustained in a setting where community attitudes depend on a longer history of a firm's actions.

In assessing the welfare implications of informal regulation, we consider the fraction of the time the firm engages in environmentally-responsible behavior. We focus on the steady-state distribution of time spent in each state, but it should be understood that underpinning this is a rich dynamic story. Steady state does not imply here that behavior is unchanging – in the steady state community attitude can be 'bouncing' backwards and forwards between periods of support and hostility, and the firm's behavior between clean and dirty.

Markov models, which allow a key state variable to vary stochastically through time, have been used in many areas of the social sciences, including economics. They have been used to model the evolution of individual tastes and behaviors (Pearson and West (2003), Netzer (2008), Biehl (2001)) and mass attitudes (Yu and Pei (2009)). In Lagunoff (2006) the policy bias of an incumbent regulator evolves through time according to a Markov process, while in Blomberg et al. (2004) the social preconditions for terrorism do. Greenberg (1984), Harrington (1988) and Landsberger and Meilijson (1982) have used it in regulatory compliance settings. Harrington (1988) is probably the paper closest in spirit to ours, with the attitude of the regulatory agency towards a particular firm varying through time in a way that is sensitive to that firm's compliance history.

The key results of our paper can be categorized threefold: (a) informal mechanisms are not as efficient as well-designed formal ones; (b) informal regulation may be good or bad for welfare; (c) informal and formal mechanisms may be substitutes or complements. The generally optimistic views that community pressure has the potential to replace formal governance, and that formal and informal incentives for regulatory compliance are necessarily additive in their contribution to welfare, need to be treated with caution.

2 Model

A firm operates in a community in each of an infinite sequence of periods. For ease of analysis we treat the community as a single entity, and assume that at any given moment its attitude towards the firm is either 'hostile' (denoted as h) or 'supportive' (denoted as s). If the community is supportive the firm's gross profit per period is $\pi(s)$, while if the

community is hostile periodic profit is $\pi(h)$, with

$$\pi(s) > \pi(h) > 0.$$

The assumption that $\pi(s) > \pi(h)$ – a supportive community is more profitable for the firm – is pivotal to the analysis. It drives the profit-motivated firm to act in a manner that avoids community hostility, providing the potential for informal regulation to work. The precise mechanism that links hostility to lower profits is black-boxed and not important for current purposes.[2]

The firm decides each period whether to behave in a manner that is 'clean' or 'dirty'. We can think of this as a model of compliance, with the binary choice corresponding to compliance or non-compliance with some legal requirement (where one exists). But the formulation allows for a more general interpretation, where firms choose to be clean because it is seen by the local community to be the 'right thing to do'.

We assume that being clean is costly, but the value of that cost is uncertain. The cost to the firm of type i of being clean is c_i per period where c_i is drawn from the distribution $F(c)$ with associated density function $f(c)$. The distribution is common knowledge, but the realization is observed privately by the firm. In contrast, it is costless to be dirty. Importantly, the firm can switch over time between these actions, clean and dirty, and does so in response to incentives.

We begin our analysis by assuming that the community is initially supportive towards the firm. If the firm chooses 'dirty' the community may be roused to hostility. Concretely, we assume that if the firm chooses dirty in period t the probability that the supportive community will be rendered hostile by the start of period $(t+1)$ is $\beta \in [0, 1]$. In contrast, the community remains supportive towards any firm that behaves in a responsible manner: that is, if the firm chooses 'clean' in period t, community support is retained with probability 1 in the next period.

The value of responsible behavior lies not merely in its ability to retain community support, but also in its redemptive effects. We assume that where the firm's previous

[2]Our preferred working assumption is that community hostility translates into a weakening of demand for the output of the firm: either a subset of consumers decide not to buy from that firm, or continue to buy but have a lower willingness to pay. We use this to motivate the welfare function that we adopt later.

action has antagonized the community, the hostility may be calmed by subsequent clean behavior. In particular if facing a hostile community the firm chooses 'clean' in period t, then the probability that the community will turn supportive in period $(t+1)$ is $\gamma \in [0,1]$. In contrast, continued dirty behavior leads hostility to persist with certainty.

The state of community feelings towards the firm can, then, be described as a Markov process, where the transition across states s and h depends on the current action chosen by the firm. The pair of parameters β and γ capture the process of community attitude formation – how ready it is to withdraw support in response to episodes of dirtiness, and how readily calmed it is when that dirtiness ceases. These parameters would depend, among other things, on how informed the community is – an alert community might be more responsive to the firm's actions, with higher values for both β and γ.

The firm is a profit maximizer (recall our earlier comments on the usage of the label 'corporate social responsibility' in these settings) and chooses its actions to maximize the present value of current and future payoffs over an infinite horizon; future values are discounted using a periodic discount factor $\delta < 1$. In general, the firm's time profile of choices can be complex. However, we rely here on the well-known ergodic theorem for such Markov decision problems: in such cases the optimal policy of the firm is stationary, in that decisions at any time depend only on the current state of system.

It is sufficient, then, to look at policies in which the firm chooses a state-dependent action. A decision at any time is a map from the set of states (supportive or hostile, s and h) to the set of actions (clean or dirty, $\mathcal{C}$ and $\mathcal{D}$). A policy refers to a rule for making decisions over time. The set of feasible policies is given by a map $\Omega : \{s \times h\} \rightarrow \{\mathcal{C} \times \mathcal{D}\}$ where a policy specifies an action for each state. With two states and two possible actions, there are four distinct policies.

Let $\Omega^{a_s a_h}$ denote the policy in which the firm chooses action a_s when the community is supportive and a_h when it is hostile. For instance, $\Omega^{\mathcal{CD}}$ denotes a policy in which the firm chooses 'clean' when the community is supportive, but 'dirty' when it is hostile. With policy $\Omega^{\mathcal{DD}}$ it picks 'dirty' regardless of the state of community attitude.

Of particular interest are $\Omega^{\mathcal{CC}}$ and $\Omega^{\mathcal{DC}}$, which involve the use of responsible behavior – choosing clean – in the face of a hostile community. Policy $\Omega^{\mathcal{CC}}$ involves responsible

behavior even when the community is supportive, and we will refer to such activity as 'retentive CSR'. Policy $\Omega^{\mathcal{DC}}$ is more opportunistic – it involves responsible behavior only when the firm has lost community support. Since it does so in order to restore community support we will refer to this policy as 'redemptive CSR'.

In what follows we evaluate the expected return to various stationary policies as a function of c_i, the cost of clean behavior. It turns out that by choosing $\Omega^{\mathcal{CD}}$ the firm can do no better than by choosing $\Omega^{\mathcal{CC}}$ or $\Omega^{\mathcal{DD}}$, so we can ignore $\Omega^{\mathcal{CD}}$ whenever both those policies are available. Hence we examine how the firm's optimal choice among the three remaining policies varies with c_i. For a particular distribution of firm types this will allow us to characterize the likelihood of socially responsible behavior, and thereby to assess the efficiency of the pattern of behavior generated by the informal regulation.

2.1 Retentive CSR

Policy $\Omega^{\mathcal{CC}}$ prompts the firm to choose 'clean' regardless of its current standing in the community. Starting with supportive community, consistent clean behavior retains support of the community, giving the firm a net profit of $[\pi(s) - c_i]$ per period indefinitely. The expected payoff to this policy is the value of that stream of payoffs, capitalized using discount factor δ.

$$E^{\mathcal{CC}}(s, c_i) = \frac{[\pi(s) - c_i]}{1 - \delta}.$$

For notational ease we define $\Pi(s) = \frac{\pi(s)}{1-\delta}$ to be the present discounted value of gross profits with a perpetually supportive community, and $C_i = \frac{c_i}{1-\delta}$ to be the present discounted value of the costs c_i incurred every period. Then, with slight abuse of notation, we can write the expected payoff to policy $\Omega^{\mathcal{CC}}$ as

$$E^{\mathcal{CC}}(s, C_i) = \Pi(s) - C_i. \tag{1}$$

2.2 Redemptive CSR

Policy $\Omega^{\mathcal{DC}}$ is more opportunistic. When the community is supportive, the firm picks 'dirty', thereby avoiding cost c_i. However if the community turns hostile the firm picks 'clean' in an effort to restore community support. In essence, the firm's CSR is purely

redemptive in intent.[3]

Once again, we evaluate the expected payoff to this policy starting with a community that is supportive. The expected payoff can be expressed recursively as

$$E^{\mathcal{DC}}(s, c_i) = \pi(s) + \delta\ [\beta E^{\mathcal{DC}}(h, c_i) + (1 - \beta)E^{\mathcal{DC}}(s, c_i)]. \tag{2}$$

To see why note that, under this policy, the firm earns $\pi(s)$ in the initial period and the outcome next period is stochastic: with probability β the community turns hostile, with continuation value $E^{\mathcal{DC}}(h, c_i)$, and with probability $(1 - \beta)$ the community remains supportive.

To compute the continuation value $E^{\mathcal{DC}}(h, c_i)$, note that this redemptive policy calls for clean behavior in response to hostility: that generates current payoff $[\pi(h) - c_i]$, with a stochastic future outcome: with probability γ the hostile community will be calmed by the firm's conversion to clean behavior, and with the residual probability it will remain hostile. We have

$$E^{\mathcal{DC}}(h, c_i) = [\pi(h) - c_i] + \delta\ [\gamma E^{\mathcal{DC}}(s, c_i) + (1 - \gamma)E^{\mathcal{DC}}(h, c_i)]. \tag{3}$$

Equations (2) and (3) can be solved together to obtain a closed-form solution. To save notation, we define $\Lambda = (1 - \delta(1 - \beta))$ and $\Psi = (1 - \delta(1 - \gamma))$, and, as before, we write $\Pi(h) = \frac{\pi(h)}{1-\delta}$. We then obtain

$$E^{\mathcal{DC}}(s, C_i) = \omega_{\mathcal{DC}}\ \Pi(s) + (1 - \omega_{\mathcal{DC}})\ [\Pi(h) - C_i], \tag{4}$$

where

$$\omega_{\mathcal{DC}} = \tfrac{(1-\delta)\Psi}{\Psi\Lambda - \delta^2\gamma\beta}. \tag{5}$$

Effectively, under policy $\Omega^{\mathcal{DC}}$ the firm transitions stochastically between the two states. When the community is supportive the firm chooses to be dirty and earns $\pi(s)$ per period. As and when dirty behavior tips the community into hostility, the firm switches to clean and earns $\pi(h) - c_i$ per period till the community becomes supportive again. The weights $\omega_{\mathcal{DC}}$ and $(1 - \omega_{\mathcal{DC}})$ capture the average time spent in the two states, given the stochastic process underlying the evolution of community attitudes.

[3]Consider, for example, BP's actions in the wake of the Deep Horizon disaster, which Parsons (2011) labels as 'reputation repair'.

2.3 No CSR

Under policy $\Omega^{\mathcal{DD}}$ the firm never engages in clean behavior, regardless of community attitude. Starting with an initially supportive community, the present value of the payoff to this policy is

$$E^{\mathcal{DD}}(s, c_i) = \pi(s) + \delta \left[\beta E^{\mathcal{DD}}(h, c_i) + (1 - \beta) E^{\mathcal{DD}}(s, c_i)\right]. \tag{6}$$

During the initial 'honeymoon' period (which could be a single period or longer) the firm is able to extract profit $\pi(s)$ per period, without any expenditure on being clean. But after that initial support has been dissipated, the firm simply accepts its unpopularity and settles for periodic payoffs $\pi(h)$ indefinitely. It is easy to see that

$$E^{\mathcal{DD}}(h, c_i) = \Pi(h). \tag{7}$$

Combining (6) and (7), we have

$$E^{\mathcal{DD}}(s, C_i) = \omega_{\mathcal{DD}} \Pi(s) + (1 - \omega_{\mathcal{DD}}) \Pi(h), \tag{8}$$

where

$$\omega_{\mathcal{DD}} = \tfrac{1-\delta}{\Lambda}. \tag{9}$$

2.4 Firm's choice of optimal policy

The firm understands the benefits to maintaining community support – that bad behavior can induce hostility but this can be assuaged by subsequent good behavior. The optimal CSR policy for the firm – we denote this as $\Omega^*(c_i, \beta, \gamma)$ – is the one that maximizes the net present value of expected returns detailed above, given the parameter values c_i, β and γ. Given the stationarity of the environment, the optimal policy is time invariant and is not sensitive to choice of initial state.[4]

Our focus is on how the firm's optimal choice depends on the realization of cost, c_i. It is easy to compare the policies if clean behavior is costless: we have

$$E^{\mathcal{CC}}(s, 0) > E^{\mathcal{DC}}(s, 0) > E^{\mathcal{DD}}(s, 0) > 0.$$

[4]This second feature ensures that our assumption that communities are initially supportive can be made without loss of generality.

If clean behavior is costless, retentive CSR policy is optimal as it avoids any hostility that damages periodic payoffs. That $E^{\mathcal{DC}}(s,0) > E^{\mathcal{DD}}(s,0)$ is equally intuitive and says that if ever the firm *were* to face a hostile community it should engage in the (costless) action to restore support.

More generally, the expected returns to the three policies are linear in c_i: while $E^{\mathcal{CC}}(s,c_i)$ and $E^{\mathcal{DC}}(s,c_i)$ are decreasing in c_i, though at different rates, $E^{\mathcal{DD}}(s,c_i)$ is invariant to changes in c_i. Using (1) and (4), it is easy to see $E^{\mathcal{DC}}(s,c_i) \geq E^{\mathcal{CC}}(s,c_i)$ if and only if

$$\omega_{\mathcal{DC}}\Pi(s) + (1-\omega_{\mathcal{DC}})[\Pi(h) - C_i] \geq \Pi(s) - C_i,$$

or, equivalently, using (5), if and only if

$$c_i \geq \left(\frac{\delta\beta}{\Psi}\right)[\pi(s) - \pi(h)] \equiv \hat{c}. \tag{10}$$

Compared to retentive CSR (policy $\Omega^{\mathcal{CC}}$), redemptive CSR (policy $\Omega^{\mathcal{DC}}$) saves on the cost of cleanliness during periods in which the community is supportive but requires the firm to accept possible intervals of community hostility as a consequence. Whether that is a profitable trade-off depends on the cost of clean behavior: if c_i exceeds the critical threshold $\hat{c}$, the expected return to the more opportunistic policy $\Omega^{\mathcal{DC}}$ is greater than the return to policy $\Omega^{\mathcal{CC}}$.

By a similar logic, using (4) and (8), we find that the expected return to policy $\Omega^{\mathcal{DC}}$ is greater than the return to policy $\Omega^{\mathcal{DD}}$ if and only if

$$c_i \leq \left(\frac{\delta\gamma}{\Lambda}\right)[\pi(s) - \pi(h)] \equiv \hat{\hat{c}}. \tag{11}$$

This is equally intuitive. Being clean constitutes an investment by the firm in improving and/or maintaining community attitudes. If the firm finds cleanliness *very* expensive it will never find such spending worthwhile, and $\hat{\hat{c}}$ defines the threshold beyond which that is the case.

Combining (10) and (11), it is clear that policy $\Omega^{\mathcal{DC}}$ is optimal for the firm if and only if c_i lies in the interval $[\hat{c}, \hat{\hat{c}}]$. Whether such an interval exists depends on the values of $\hat{c}$ and $\hat{\hat{c}}$. Comparing (γ/Λ) and (β/Ψ), as defined earlier, leads to the following.

Remark 1 *$\hat{c} < \hat{\hat{c}}$ if and only if $\beta < \gamma$.*

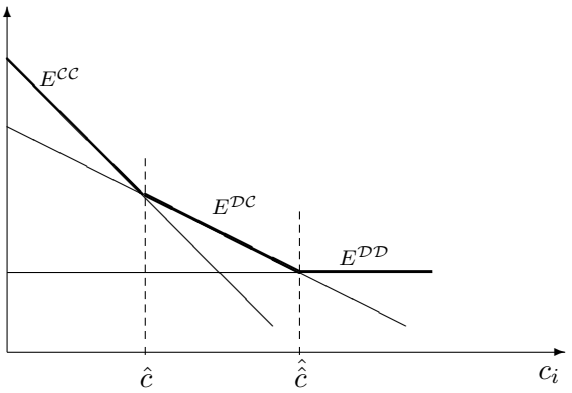

Figure 1: Optimal policy as function of c_i when $\beta < \gamma$

The restriction that $\beta < \gamma$ requires that the community is sufficiently 'forgiving': that clean behavior be more likely to restore community support than dirty behavior is to trigger hostility. If $\beta = \gamma$, we have $\hat{c} = \hat{\hat{c}}$.

Figure 1 plots the profitability of the three policies for alternative values of c_i, for the case where $\beta < \gamma$. For c_i below the lower threshold $\hat{c}$ the retentive CSR policy $\Omega^{\mathcal{CC}}$ is most profitable; above the higher threshold $\hat{\hat{c}}$ the no CSR policy $\Omega^{\mathcal{DD}}$ dominates. For the intervening range, the redemptive CSR policy $\Omega^{\mathcal{DC}}$ is best. Given a sufficiently dispersed distribution $F(c_i)$ of costs, there is an interval of low realizations for which the firm chooses to be clean at all times and an interval of high realizations for which it will never choose clean. As long as $\hat{c} < \hat{\hat{c}}$, there exists a central interval of mid-range cost realizations for which the firm chooses clean behavior only when the community is hostile. We formalize this as:

Proposition 1 *Let $\beta < \gamma$ and let $\hat{c}$ and $\hat{\hat{c}}$ be as defined above. The optimal CSR policy for the firm is*

$$\begin{array}{lll} \Omega^{\mathcal{CC}} & \text{if} \quad c_i \leq \hat{c} & \textit{Retentive CSR} \\ \Omega^{\mathcal{DC}} & \text{if} \quad \hat{c} < c_i < \hat{\hat{c}} & \textit{Redemptive CSR} \\ \Omega^{\mathcal{DD}} & \text{if} \quad \hat{\hat{c}} \leq c_i & \textit{No CSR} \end{array}$$

How does the optimal policy vary with c_i when $\beta > \gamma$? Adapting the previous arguments, we can show that for this case policy $\Omega^{\mathcal{DC}}$ is never optimal, so the choice is essentially between $\Omega^{\mathcal{CC}}$ and $\Omega^{\mathcal{DD}}$. Further $E^{\mathcal{CC}}(s, c_i) \geq E^{\mathcal{DD}}(s, c_i)$ if and only if $c_i \leq \tilde{c}$, where

$$\tilde{c} \equiv \left(\frac{\delta\beta}{\Lambda}\right)[\pi(s) - \pi(h)]. \tag{12}$$

In this case the optimal policy switches from $\Omega^{\mathcal{CC}}$ for c_i below the threshold $\tilde{c}$, to $\Omega^{\mathcal{DD}}$ above that threshold, without any intervening range where $\Omega^{\mathcal{DC}}$ is optimal.

Proposition 2 *Let $\beta > \gamma$ and let $\tilde{c}$ be as defined above. The optimal CSR policy for the firm is*

$$\begin{aligned} \Omega^{\mathcal{CC}} \quad & \text{if} \quad c_i \leq \tilde{c} \quad \textit{Retentive CSR} \\ \Omega^{\mathcal{DD}} \quad & \text{if} \quad c_i > \tilde{c} \quad \textit{No CSR} \end{aligned}$$

We later explore the implications of this case in our welfare analysis.

2.5 Steady state

There is a potential divergence between the pay-offs to an optimal policy during the initial 'honeymoon' period (associated with the assumption of an initially supportive community) and the flow of payoffs in steady state. In what follows we move to steady-state considerations. For any chosen policy Ω, the steady state under the Markov process would typically imply continuous transitions across states, with a limiting distribution of average time spent in each state. Some policies may result in 'absorbing states' but in general steady state does not imply that behavior is unchanging – in a steady state community attitude can be 'bouncing' backwards and forwards between periods of support and hostility, and the firm's behavior between clean and dirty.

For any policy Ω and the associated steady state distribution, we define $p(\Omega)$ to be the fraction of periods in which the firm ends up choosing 'clean'. The evaluation of $p(\Omega)$ is straightforward for some policies. In particular, policy $\Omega^{\mathcal{CC}}$ requires the firm to be always clean so $p(\Omega^{\mathcal{CC}}) = 1$ (with s being an 'absorbing state'). Under policy $\Omega^{\mathcal{DD}}$ we have $p(\Omega^{\mathcal{DD}}) = 0$.

Consider the case in which the firm adopts policy $\Omega^{\mathcal{DC}}$, which calls for clean behavior only when the community is hostile. In steady state the fraction of time that the firm spends in each state is given by the limiting distribution of a Markov process with the following transition matrix:

$$\begin{bmatrix} (1-\beta) & \beta \\ \gamma & (1-\gamma) \end{bmatrix}.$$

It is easy to verify that

$$p(\Omega^{\mathcal{DC}}) = \frac{\beta}{\beta + \gamma},$$

the expected steady-state fraction of time in which the community is hostile under this policy. The firm is dirty for the residual fraction of time $\frac{\gamma}{\gamma+\beta}$, when it faces a supportive community. The expected steady-state return to policy $\Omega^{\mathcal{DC}}$ is

$$E^{\mathcal{DC}}(C_i) = \omega^*_{\mathcal{DC}}\Pi(s) + (1 - \omega^*_{\mathcal{DC}})[\Pi(h) - C_i], \tag{13}$$

where

$$\omega^*_{\mathcal{DC}} = \frac{\gamma}{\gamma + \beta}. \tag{14}$$

In contrast, the steady-state returns to policies $\Omega^{\mathcal{CC}}$ and $\Omega^{\mathcal{DD}}$ are

$$E^{\mathcal{CC}}(C_i) = \Pi(s) - C_i, \tag{15}$$

and

$$E^{\mathcal{DD}}(C_i) = \Pi(h). \tag{16}$$

Propositions 1 and 2 had characterized the firm's optimal policy given an initially supportive population. The steady-state versions of those results are analogous. Replicating our previous steps, we define the lower and upper cost thresholds in order to characterize the optimal policy:

$$c_*(\beta, \gamma) = \left(\frac{\beta}{\gamma}\right)[\pi(s) - \pi(h)], \tag{17}$$

and

$$c^*(\beta, \gamma) = \left(\frac{\gamma}{\beta}\right)[\pi(s) - \pi(h)]. \tag{18}$$

Proposition 3 *Let $\beta < \gamma$ and let c_* and c^* be as defined above. The optimal steady-state CSR policy for the firm given cost realization c_i is*

$$\begin{array}{llll} \Omega^{\mathcal{CC}} & \textit{if} & c_i \leq c_* & \textit{Retentive CSR} \\ \Omega^{\mathcal{DC}} & \textit{if} & c_* < c_i < c^* & \textit{Redemptive CSR} \\ \Omega^{\mathcal{DD}} & \textit{if} & c^* \leq c_i & \textit{No CSR} \end{array}$$

Figure 2 plots the expected amount of time that the firm spends engaged in clean conduct, in steady state, given its realization of c_i. As before Ω^* denotes the optimal policy. For low cost realizations ($c_i < c_*$) the firm finds it optimal to choose $\Omega^{\mathcal{CC}}$, so always chooses clean behavior: we have $p(\Omega^* |\ c_i < c_*) = p(\Omega^{\mathcal{CC}}) = 1$. For high realizations

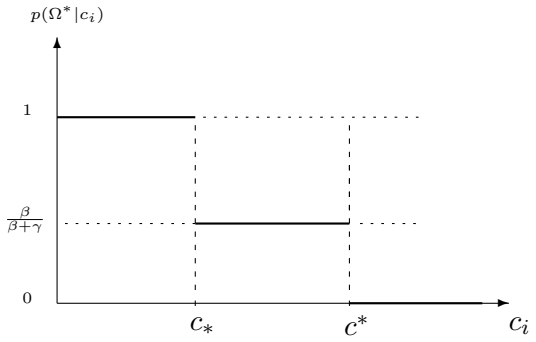

Figure 2: Propensity to clean conduct as function of c_i, when $\beta < \gamma$

($c_i > c^*$) it is optimal to choose $\Omega^{\mathcal{DD}}$, so the firm is never clean. In the intervening range the firm is clean for a fraction $p(\Omega^{\mathcal{DC}}) = \frac{\beta}{\beta+\gamma}$ of the time.

For any distribution $F(c)$, we can compute the ex-ante probability of clean conduct.[5]

Remark 2 *Let $\beta < \gamma$. Given cost distribution $F(c)$, in steady state the expected fraction of time that a firm chooses clean behavior is*

$$\int p(\Omega^*|c_i)f(c)dc = \int_0^{c_*} f(c)dc + \int_{c_*}^{c^*} \left(\frac{\beta}{\beta+\gamma}\right) f(c)dc.$$

The values obtained earlier for the thresholds permit a more concise expression. Write $\overline{\pi} = \pi(s) - \pi(h)$, so $c_* = \frac{\beta}{\gamma}\overline{\pi}$ and $c^* = \frac{\gamma}{\beta}\overline{\pi}$. Then, the expected fraction of time the firm expects to be clean is

$$F(c_*) + \left(\tfrac{\beta}{\beta+\gamma}\right)[F(c^*) - F(c_*)] = \tfrac{\gamma}{\beta+\gamma}F\left(\tfrac{\beta}{\gamma}\overline{\pi}\right) + \tfrac{\beta}{\beta+\gamma}F\left(\tfrac{\gamma}{\beta}\overline{\pi}\right).$$

What if $\beta > \gamma$? Here the redemptive CSR policy $\Omega^{\mathcal{DC}}$ is never optimal. As c_i varies, the optimal policy switches from $\Omega^{\mathcal{CC}}$ to $\Omega^{\mathcal{DD}}$, with the critical threshold being $\overline{\pi} = \pi(s) - \pi(h)$.

Proposition 4 *Let $\beta \geq \gamma$. The optimal steady-state CSR policy for firm i in steady state is*

$$\begin{array}{lll} \Omega^{\mathcal{CC}} & \text{if} \quad c_i \leq \overline{\pi} & \textit{Retentive CSR} \\ \Omega^{\mathcal{DD}} & \text{if} \quad c_i > \overline{\pi} & \textit{No CSR} \end{array}$$

Remark 3 *If $\beta \geq \gamma$, in steady state the fraction of time that the firm expects to be engaged in clean behavior is*

$$\int p(\Omega^*|c_i)f(c)dc = \int_0^{\overline{\pi}} f(c)dc = F(\overline{\pi}).$$

[5]Note that the current model is presented for a single firm located in a single community. If however we extend to think of a population of communities, embedded in each is a single firm, the expression becomes the expected fraction of firms engaged in good behavior in a given period.

In what follows, we focus our attention on the case where $\beta < \gamma$ taking care to identify, at relevant places, the implications of departing from this assumption.

3 Community pressure and welfare

Having explored the optimal choice of CSR policy for the firm, we turn next to normative questions. Is community hostility welfare-improving? How close is the resulting set of incentives to the first-best outcome?

We focus on the properties of the steady state outcome for our welfare analysis. This approach, quite usual in models of this sort, has the advantage of abstracting from any bias due to arbitrary choice of initial state.

In developing a welfare function there is the issue of treatment of the profit 'penalty' that the community imposes by withdrawing its support, namely $\overline{\pi} = \pi(s) - \pi(h)$. In particular, does that penalty impose a real resource burden upon welfare, or is simply a *transfer* away from the reference firm? The appropriate weight on the penalty in social welfare might vary with the channel through which the community vents its hostility. If members of a hostile local community pick up rocks and throw them through factory windows, then the penalty can reasonably be considered a welfare burden. However, if members of that community decide not to buy services from the firm that has aroused hostility, but rather take their custom to some other firm in the same economy, then the penalty is merely a transfer and should not appear as a net burden on welfare. In general, we could introduce a coefficient $k \in [0, 1]$ to capture the fraction of penalty $\overline{\pi}$ that is a real resource burden in the social welfare function. However, for simplicity we implicitly set that parameter k equal to zero and exclude the penalty term from welfare comparisons. Nothing critical rests upon this simplification.

For exposition we will examine departures from the first-best outcome. Let b denote the per period social benefits of clean behavior. This is the environmental damage avoided when the firm is clean. First-best CSR requires firm i choosing clean every period if $c_i < b$, and dirty otherwise. Put simply, the firm should choose clean if and only if the private (= social) cost is less than the social benefit.

We define the expected social loss function in terms of deviations from this first-best:

$$SL = \int_0^b (1 - p(\Omega^*(c_i))(b - c_i)f(c)dc + \int_b^\infty p(\Omega^*(c_i))(c_i - b)f(c)dc,$$

where the optimal policy achieved through informal regulation, $\Omega^*(c_i; \beta, \gamma)$, varies with the configuration of β and γ. The first composite term captures the loss associated with a low-cost firm (one with $c_i < b$) not doing enough CSR relative to the first-best outcome, and the second with a high-cost firm ($c_i > b$) doing too much.

We begin our analysis with the case where $\beta < \gamma$, so that the optimal policy for the firm is, as in Proposition 3, determined by its cost c_i relative to thresholds c_* and c^*. We consider three distinct sub-cases, based on where b lies relative to the interval $[c_*, c^*]$. Our model makes no *a priori* assumption about the magnitude of b, the social benefit of clean behavior, relative to the communal penalty $\overline{\pi}$ which affects the thresholds c_* and c^*: if so, b could lie anywhere relative to this interval.

Consider, first, the sub-case where $c_* < b < c^*$. If $c_i \leq c_*$ the firm chooses a policy that entails clean behavior at all times, consistent with first-best for this case. For $c_i \geq c^*$ the firm chooses a policy that entails dirty behavior at all time, once again consistent with first-best. For realizations $c_* < c_i < c^*$ the firm chooses policy $\Omega^{\mathcal{DC}}$, which entails clean behavior for a fraction $\frac{\beta}{\gamma+\beta}$ of the time. Relative to the first-best, this is not enough for $c_i \in (c_*, b)$ and excessive for realizations $c_i \in (b, c^*)$. The evaluation of expected social loss in this case is, then,

$$SL_{|[c_*<b<c^*]} = \int_{c_*}^b \left(\tfrac{\gamma}{\beta+\gamma}\right)(b - c_i)f(c)dc + \int_b^{c^*} \left(\tfrac{\beta}{\beta+\gamma}\right)(c_i - b)f(c)dc. \tag{19}$$

The welfare loss is easily understood from Figure 3. The bold line indicates the fraction of the time the firm engages in clean behavior as a function of cost type, given our regime of informal regulation. In contrast, the dotted lines denotes the first-best response given b.

Consider, next, the more extreme sub-case where $b \leq c_*$. With relatively low social benefit, clean behavior is not socially optimal for most cost realizations. Nonetheless the threat of community hostility drives the firm to clean behavior too often. We have, for this case

$$SL_{|[b<c_*]} = \int_b^{c_*} (c_i - b)f(c)dc + \int_{c_*}^{c^*} \left(\tfrac{\beta}{\beta+\gamma}\right)(c_i - b)f(c)dc. \tag{20}$$

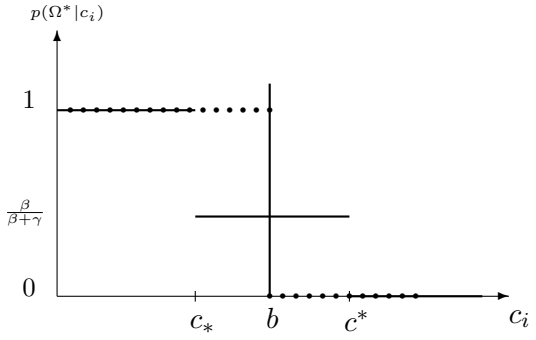

Figure 3: Social loss associated with CSR: case where $\beta < \gamma$ and $c_* < b < c^*$

At the other extreme, if $b \geq c^*$, the community penalty is so weak that clean behavior is not induced often enough, with expected social loss

$$SL_{|[b>c^*]} = \int_{c_*}^{c^*} \left(\frac{\gamma}{\beta+\gamma}\right)(b - c_i)f(c)dc + \int_{c^*}^{b}(b - c_i)f(c)dc. \tag{21}$$

Regardless of which sub-case is relevant, we have the following proposition.

Proposition 5 *If $\beta < \gamma$, the expected pattern of CSR induced by the threat of community hostility is not first-best.*

To understand this note that if $\beta < \gamma$, as Figure 3 indicates, the firm's CSR policy response involves a step function $p(\Omega^*)$ with *two* downward steps as c_i varies from low to high values. On the other hand, the first-best policy – call it Ω^{FB} – calls for a single step

$$p(\Omega^{FB}) = \begin{cases} 1 & \text{if } c_i < b, \\ 0 & \text{otherwise.} \end{cases}$$

If so, community pressure is inefficient.

Is community pressure more efficient if $\beta > \gamma$? Recall, from Proposition 4, that for this case the firm chooses policy $\Omega^{\mathcal{CC}}$ for $c_i \leq \overline{\pi}$ and $\Omega^{\mathcal{DD}}$ otherwise. Even in this case first-best would be achieved through the threat of community hostility only if the penalty $\overline{\pi}$ happens to equal precisely b – a measure zero event. Generically, even when $\beta > \gamma$, first-best is not achieved.

This generic inefficiency of informal regulation when $\beta > \gamma$ is significant. While our model uses the simplification that transition probabilities β and γ are fixed parameters, in reality they may well depend on b, the damage associated with dirty behavior. The community may turn hostile more readily when the damage is large, so that $\beta(b)$ may well

be an increasing function. If so, it is more likely that $\beta(b)$ exceeds γ when b is relatively large but, as our argument above shows, even then informal regulation will not achieve first best.

In general, the divergence may be so significant that pressure-induced CSR may be sub-optimal, not only in relation to the first-best, but relative to an outcome that involves no CSR at all.

Proposition 6 *Clean behavior generated by the threat of community hostility may lead to higher or lower welfare than would be the case if there was no clean behavior.*

To see why, note that net benefit of community pressure induced CSR activity, relative to the case where there is no restraint on firm behavior, is

$$\int p(\Omega^*(c_i))(b - c_i)f(c)dc. \tag{22}$$

Consider a situation where b is very small, or even non-existent; that is, assume b tends to zero. If strong communal penalty $\overline{\pi}$ results in $p(\Omega^*(c_i)) > 0$ for a significant range of c_i, CSR is welfare-reducing. On the other hand, with large b, some CSR is better than none. In general, the effect of informal regulation on welfare is ambiguous.

This ambiguity can be illustrated with a simple numerical example. Suppose community hostility hurts profitability by one unit (that is, $\overline{\pi} = 1$); let $\beta = 0.4$ and $\gamma = 0.6$, so that $c_* = 0.67$ and $c^* = 1.5$. Assume c is distributed uniformly in the interval $[0, 2]$, so that $F(c) = 0.5c$ for $0 \leq c \leq 2$. The optimal CSR response to informal regulation entails the firm choosing clean behavior with probability one if its cost realization $c_i \leq 0.67$, and with probability zero for $c_i > 1.5$. For costs in the intervening range, the firm's optimal policy entails clean behavior with probability 0.6 in steady state. Whether clean behavior induced by such community pressure is welfare improving or not depends on b, the social benefit of clean behavior. It is easy to check that if b is relatively low – less than 0.58 for the above parameter configuration – CSR is welfare reducing. Values of b larger than this value make CSR desirable.

The possibility that community hostility may drive firms to behave in a manner that has real costs but limited social benefits is consistent with the empirical finding of Dasgupta

et al. (2000), who found that community pressure – what they refer to as 'extra legal factors' – can induce over-compliance with regulatory standards and an inefficient pattern of abatement decisions.

4 Extension: incorporating history

Our assumption that community attitude evolves solely in response to the current actions of the firm seems restrictive. Popular sentiment is often quite sensitive to the observed *history* of firms' behavior. For instance, it may be reasonable to suppose that a community is more likely to turn hostile towards a firm whose current dirty action is a continuation of a prior trend of bad behavior (a 'persistent offender') than when it comes against a more favorable record. How can our analysis accommodate such richer settings?

Observe that our purpose here is not a complete characterization of the firm's rational policy response to informal regulation, but rather to highlight the fact that these responses will typically involve departures from what is socially optimal. Recall that the first-best requires that the firm choose clean if $c_i < b$, and dirty otherwise. This outcome does not obtain – except by contrivance – in our base model: informal regulation is (almost surely) inefficient. In this section we adapt our model to allow the possibility that the evolution of community attitudes may be shaped by a firm's past behavior in addition to its current action, to confirm that our basic insight is robust to this extension.

One natural way to incorporate history dependence is to enlarge the set of states to include the firm's recent environmental record. We continue to assume that community attitude in any period is binary – supportive or hostile. However we can extend the specification of a typical state at time t by including the history $\mathcal{H}_{t-1}$ of the firm's past behavior: its record of clean or dirty choices in previous periods. For instance, $(s, \mathcal{H}_{t-1})$ specifies a state in which the community is supportive and recalls the firm's environmental history $\mathcal{H}_{t-1}$. Transitions between supportive and hostile communities would now depend not just upon current actions, but also on the firm's record of past behavior. So the probability that dirty behavior renders a supportive community hostile, for example, can now be made sensitive to how the firm behaved in the past.

The enlarged set of states supports a much richer set of policies (recall that a policy specifies an action for the firm – clean or dirty – for every distinct state). Any chosen policy implies a pattern of transitions across states, based on assumed transition probabilities. Some policies may lead to absorbing states. If, for instance, a policy of 'always clean' (that is, a policy of clean regardless of current state) preserves community support with certainty, it would lead to an absorbing state with payoff $\pi(s) - c_i$ every period. On the other hand, a policy of 'always dirty' would typically result in an absorbing state of community hostility, with payoff $\pi(h)$ per period. When c_i, the cost of being clean, is relatively low the former policy would be preferable; when it is relatively high, the latter would.

But there are intermediate ranges of c_i in which other policies might dominate the 'always clean' and 'always dirty' policies. Firms realize that in a dynamic setting clean behavior has the potential to preserve (or restore) community support, allowing it to earn higher profits. At the same time a firm facing a supportive community might be tempted to behave opportunistically, avoiding the cost of being clean especially if it believes that such lapses will not necessarily compromise community support.

Without loss of generality, assume that some arbitrary policy Ω involves clean behavior for a fraction $p(\Omega)$ of the periods and is expected to retain community support for a fraction $x(\Omega)$ of the time. The expected payoff to this policy is

$$E^{\Omega}(C_i) = x(\Omega)\Pi(s) + [1 - x(\Omega)]\Pi(h) - p(\Omega)C_i. \tag{23}$$

Such a policy will dominate a policy of 'always clean' if

$$x(\Omega)\pi(s) + [1 - x(\Omega)]\pi(h) - p(\Omega)c_i \geq \pi(s) - c_i, \tag{24}$$

or equivalently if

$$c_i \geq \frac{1 - x(\Omega)}{1 - p(\Omega)}[\pi(s) - \pi(h)] \equiv c_*(\Omega). \tag{25}$$

At the same time this policy will dominate 'always dirty' if

$$x(\Omega)\pi(s) + [1 - x(\Omega)]\pi(h) - p(\Omega)c_i \geq \pi(h), \tag{26}$$

or equivalently if

$$c_i \leq \frac{x(\Omega)}{p(\Omega)}[\pi(s) - \pi(h)] \equiv c^*(\Omega). \tag{27}$$

The range $[c_*, c^*]$ is non-trivial (that is, we have $c^*(\Omega) > c_*(\Omega)$) whenever $\frac{x(\Omega)}{p(\Omega)} > \frac{1-x(\Omega)}{1-p(\Omega)}$, or equivalently, whenever $x(\Omega) > p(\Omega)$. This condition has an easy interpretation – such opportunistic policies dominate *both* 'always clean' and 'always dirty' for some intermediate values of c_i whenever episodes of clean behavior can deliver community support for disproportionately long periods.[6]

The existence of such an intermediate range of c_i where the firm chooses to be clean for only a fraction of the time implies departures from social optimality which, recall, requires that the firm choose clean whenever $c_i < b$ and remain dirty otherwise.

To illustrate this argument consider a setting in which community attitudes depend on the firm's most recent choice and its choice in the previous period (in other words on its two-period history). Limiting community recall to only one previous period allows us some tractability: longer periods of recall can be handled with additional complexity. With one-period recall, the typical state in any period is given by the conjunction of the current attitude of the community and its observed history of the firm's choice in the previous period. For instance, state $(s, \mathcal{D}_{t-1})$ at time t describes a community that is supportive despite its recollection that the firm had been dirty in the previous period. The set of possible states at time t is then:

$$\{(s, \mathcal{C}_{t-1}), (s, \mathcal{D}_{t-1}), (h, \mathcal{C}_{t-1}), (h, \mathcal{D}_{t-1})\}.$$

As before the evolution of community feelings is given by a Markov process, which describes the transitions across the four history-contingent states. In the spirit of our previous setting, we assume that if the community is already supportive clean behavior in the current period preserves that support. Similarly, if the community is already hostile, dirty behavior in the current period preserves that hostility.

However, to allow community attitudes to be shaped by the past record of the firm we assume that dirty behavior is more likely to render a supportive community hostile

[6]Proposition 3 is a special case of this result. Without history-dependence, under policy $\mathcal{DC}$, the community is supportive (and the firm dirty in response) for fraction $\frac{\gamma}{\beta+\gamma}$ of the time, while the community is hostile (and firm clean) for fraction $\frac{\beta}{\beta+\gamma}$. So $x(\mathcal{DC}) = \frac{\gamma}{\beta+\gamma}$ and $p(\mathcal{DC}) = \frac{\beta}{\beta+\gamma}$. Whenever $x(\mathcal{DC}) > p(\mathcal{DC})$ or, equivalently $\gamma > \beta$, there exists a range of c_i for which policy $\mathcal{DC}$ dominates both $\mathcal{CC}$ and $\mathcal{DD}$.

Further, the underlying argument holds for any policy, including non-steady-state ones. Focusing on steady state policy provide us the tractability to evaluate $x(\Omega)$ and $p(\Omega)$.

if there is previous history of dirty behavior. Specifically, let β_1 be the probability that repeated dirty behavior $\mathcal{D}_{t-1}\mathcal{D}_t$ triggers hostility, while β_0 is the corresponding probability for $\mathcal{C}_{t-1}\mathcal{D}_t$. We assume that $\beta_1 > \beta_0$.

Similarly we assume clean behavior is more likely to restore community support if there is a track record of clean behavior. Specifically, let γ_0 be the probability that repeated clean behavior $\mathcal{C}_{t-1}\mathcal{C}_t$ restores support, while γ_1 is the corresponding probability for $\mathcal{D}_{t-1}\mathcal{C}_t$. We assume that $\gamma_0 > \gamma_1$.

A policy specifies an action for every state, where states are listed in the order specified above. For instance, under policy $\mathcal{CCCC}$ the firm chooses clean in each of the four possible state. Under the more opportunistic policy $\mathcal{DDCC}$ it chooses dirty for states in which the community is supportive and clean whenever it is hostile. The richer set of history-contingent state allows for more nuanced policies: for instance, the policy $\mathcal{DDDC}$ is even more opportunistic, requiring clean behavior only in the face of a hostile community that remembers previous dirty action.

With four states and two actions in every state, there are 16 distinct policies. As before we could compare the steady-state payoffs across all these policies but our analysis is eased by the fact that some policies are payoff-equivalent, allowing us to focus on a smaller selection of policies.

The payoff to some policies is straightforward to evaluate. The 'unconditionally clean' policy $\mathcal{CCCC}$ preserves community support forever and has a net payoff of $\Pi(s) - C_i$. The 'unconditionally dirty' policy $\mathcal{DDDD}$ has a net payoff of $\Pi(h)$. The steady-state payoffs associated with other policies require more careful calculation. For instance, the opportunistic policy $\mathcal{DCCC}$, which calls for clean behavior in every state except $(s, \mathcal{C}_{t-1})$ will involve repeated transitions between various states. Replicating arguments along previous lines, we can show that this opportunistic policy dominates both the 'unconditionally clean' policy $\mathcal{CCCC}$ and 'unconditionally dirty' policy $\mathcal{DDDD}$ for some range of costs.[7]

Of course, that some policy dominates both $\mathcal{CCCC}$ and $\mathcal{DDDD}$ for some range of costs does not imply that it is necessarily the best among all available policies. The firm's

[7]Specifically, policy $\mathcal{DCCC}$ dominates policies $\mathcal{CCCC}$ and $\mathcal{DDDD}$ when c_i lies in interval $[c_*, c^*]$, where $c_* = \frac{\beta_0(1+\gamma_0-\gamma_1)}{\gamma_0}\overline{\pi}$ and $c^* = \frac{\gamma_0(2-\beta_0)}{\gamma_0+\beta_0(1-\gamma_1)}\overline{\pi}$. Here $c^* > c_*$ if and only if $\gamma_0(1-\beta_0) > \beta_0(1-\gamma_1)$.

Range of costs	Optimal policy, Ω	Fraction of time clean, $p(\Omega)$
$0 < c_i \leq 0.55$	$\mathcal{CCCC}$	1
$0.55 < c_i \leq 1$	$\mathcal{DCCC}$	0.56
$1 < c_i \leq 1.5$	$\mathcal{DDCC}$	0.4
$1.5 < c_i$	$\mathcal{DDDD}$	0

Table 1: Optimal CSR policy and preponderance of clean behavior

optimal CSR strategy might involve selection of different policies for different ranges of the cost parameter. A complete analytical characterization of the firm's optimal policy is neither straightforward, nor particularly illuminating, but a numerical illustration might help.

Let $\gamma_0 = 0.6$, $\gamma_1 = 0.5$, $\beta_1 = 0.4$ and $\beta_0 = 0.3$. Assume that $[\pi(s) - \pi(h)] = 1$, so that cost of being clean c_i is measured relative to increment in profitability from a supportive rather than hostile community. Given these values by evaluating the expected return to various policies in different ranges, the optimal CSR policy for the firm contingent on its cost realization c_i is as in Table 1.

In this example we find that the optimal CSR policy involves three discrete 'steps' (in terms of likelihood of clean behavior) at critical values that depend on the transition probability parameters. In contrast, recall that the socially-optimal policy calls for a single step at b. The same qualitative insights obtain here as in the basic version presented in Section 3: If there are multiple steps then first-best is necessarily compromised; if there is a single step then that step must coincide with equality of $\overline{\pi}$ and b, which would require a particular set of conditions to hold by chance.

5 Extension: Taxation

How does the existence of community pressure articulate with the desirability and efficacy of formal regulatory instruments such as environmental taxation? Assume that the firm's choices – clean or dirty – can be observed directly by a regulator who can choose to tax dirty behavior at rate $t \geq 0$. We will assume that tax is purely redistributive so does not

feature in welfare except through impact on the firm's choice.[8]

Proposition 7 *Absent the threat of community hostility, first-best can be implemented with a regime that imposes tax $t = b$ on the action 'dirty'.*

To see why note that if community attitudes do not affect firm profitability, the penalty $\pi(s) - \pi(h)$ disappears. Any deterrent effect will then come from taxation alone. In any period, a firm will choose clean if and only if $t \geq c_i$. If t is set equal to b, the firm will choose clean if and only if $b \geq c_i$, which corresponds to first-best. This is simply a Pigovian tax or charge on the externality that flows from the action 'dirty'.

This provides a useful benchmark. If taxes are available as an enforcement mechanism, they can achieve efficient policy management. First-best outcomes can be implemented with a straightforward Pigovian tax. We have already established (Proposition 5) that CSR is generically inconsistent with achievement of first-best outcomes, so if the choice is between a self-standing, well-designed tax regime and a system of incentives based on community hostility, the former is necessarily preferred in our setting.

The welfare analysis in Section 3 was in terms of deviations from first-best, but it can equally be interpreted as deviation from what could be achieved by judicious use of taxation.

5.1 Taxation and community pressure

What is the optimal environmental tax given the existence of community pressure? In most applied settings formal and informal incentives coexist (see, for example, Wang (2000), Pargal and Wheeler (1996)). Disclosure programs (designed to leverage public pressure for enhanced performance) will usually be introduced in settings where there is at least some existing formal policy (perhaps a tax). It is, therefore, interesting to explore how the two mechanisms articulate.

In the model here we have made no attempt to provide a micro-foundation for how

[8] In other words the marginal value of social funds is 1, in effect abstracting from so-called 'double dividend' considerations. This could be varied, to make revenue-raising an end in its own right, enhancing the attractiveness of taxation.

hostility is aroused and assuaged. Once we introduce other formal instruments, however, we might wish to revisit our assumptions about how public attitudes are formed. In a different setting Gneezy and Rustichini (2000) present experimental evidence that the introduction of a tax (or fine) can legitimize behavior otherwise seen as anti-social. In our setting, for example, local residents may be less likely to get angry with local plants that spew contaminants into the local environment if those emissions have been 'legitimized' by being taxed.

In what follows we extend our basic (memory-less) model by assuming that dirty behavior in any period incurs a unit tax t. We allow for the possibility that community hostility towards polluters is softened by taxation. If taxation 'legitimizes' dirty behavior it might lower the likelihood of triggering hostility. Specifically, we assume that the transition probability β is a decreasing function of the tax rate t, so that $\beta'(t) < 0$ for all $t \geq 0$. For the sake of tractability assume all other parameters are fixed. Taxation affects the expected return to different policies so we should expect the firm's choice of CSR policy to be sensitive to the tax rate, with optimal policy now denoted as $\Omega^*(c_i, t)$.

As before we focus on cases where $\beta(t) < \gamma$, for which the optimal CSR policy involves switching at cost thresholds, now labeled as $c_*(t)$ and $c^*(t)$.[9] To determine these thresholds, note first that $E^{\mathcal{CC}}$ is invariant to t: as policy $\Omega^{\mathcal{CC}}$ never results in dirty behavior, taxes are never paid. $E^{\mathcal{DC}}$, however, is modified by taxation: we have

$$E^{\mathcal{DC}}(C_i, t) = \omega^*_{\mathcal{DC}}(t)\left(\Pi(s) - \tfrac{t}{1-\delta}\right) + (1 - \omega^*_{\mathcal{DC}}(t))(\Pi(h) - C_i). \tag{24}$$

Higher tax reduces the firm's payoff whenever it chooses dirty and also increases the proportion $\omega^*_{\mathcal{DC}}(t) = \frac{\gamma}{\beta(t)+\gamma}$ of time spent in the supportive state by lowering $\beta(t)$, the probability of transition to the hostile state. We find that $E^{\mathcal{DC}}(c_i, t) > E^{\mathcal{CC}}(c_i, t)$ if and only if c_i exceeds the tax-modified threshold $c_*(t)$, where

$$c_*(t) = \left(\tfrac{\beta(t)}{\gamma}\right)[\pi(s) - \pi(h)] + t \leq c_*(0) + t. \tag{25}$$

Here $c_*(0)$ is the value of this threshold when the tax rate is zero. The introduction of a tax raises the lower threshold by an amount less than the tax. Similar arguments show

[9]The other case, where $\beta > \gamma$ is straightforward. Here the rational policy for the firm is to choose $\Omega^{\mathcal{CC}}$ for $c_i < \overline{\pi} + t$, and $\Omega^{\mathcal{DD}}$ otherwise. The optimal tax rate equals $b - \overline{\pi}$ in this situation, that is, at a level that corrects for the insufficiency of communal penalty relative to social benefits of clean behavior.

that taxation pushes up the upper threshold too, but by an amount greater than tax t:

$$c^*(t) = \left(\frac{\gamma}{\beta(t)}\right)[\pi(s) - \pi(h)] + t \geq c^*(0) + t. \tag{26}$$

Note that if $\beta(t)$ does not vary with t, both thresholds rise precisely by the tax t.

We analyze how the social loss, relative to the first-best, varies with the level of taxation. We evaluate the variation in social loss in the neighborhood of $t = 0$. This allows us to assess whether, starting from a regime of CSR without taxation, whether or not the introduction of a tax improves welfare.

Inserting t as an argument in the expression for social loss, $SL(t)$, we consider three subcases that differ in the magnitude of b relative to the thresholds $c_*(t)$ and $c^*(t)$. Suppose, first, that $b > c^*(0)$. Recall that here informal regulation does not induce enough clean behavior. We modify the relevant expression for social loss – see (21) – to incorporate the dependence of the thresholds on t, as in (25) and (26). Then differentiating with respect to t and using the facts that $0 < \frac{\partial c_*(t)}{\partial t} \leq 1 \leq \frac{\partial c^*(t)}{\partial t}$, we have

$$\begin{aligned}\frac{dSL(t)}{dt}_{|t=0,\ b>c^*} = &-\left[\frac{\gamma}{\beta+\gamma}(b-c_*)f(c_*)\frac{dc_*}{dt} + \frac{\beta}{\beta+\gamma}(b-c^*)f(c^*)\frac{dc^*}{dt}\right] \\ &- \frac{\gamma\beta'(t)}{(\beta+\gamma)^2}\int_{c_*}^{c^*}(b-c_i)f(c)dc.\end{aligned} \tag{27}$$

If taxation does not affect community reactions, that is if $\beta'(t) = 0$, social loss is unambiguously decreasing in t in the neighborhood of $t = 0$. The optimal tax rate is positive. However, when $\beta'(t) < 0$, it could overturn this recommendation. Imagine a situation where a minuscule tax robs informal regulation of all of its potency. Then it may be preferable, in welfare terms, to have no taxation than to dilute the deterrent effect of community pressure.

This welfare ambiguity of taxation can be illustrated with a simple numerical example. Consistent with our previous example, assume that c is distributed uniformly in the interval $[0, 2]$; that $\overline{\pi} = 1$ and $\gamma = 0.6$. Assume now that $\beta(t) = 0.4e^{-\rho t}$: here ρ determines the sensitivity of the community hostility parameter to taxation t. Let $b = 2$, so that as required for this case, we have $b > c^*(0)$. It is straightforward to check that the optimal tax rate is positive when $\beta(t)$ is not very sensitive to t (that is, ρ is small or zero). However, if ρ is larger than, say, 4, $SL(t)$ is increasing in t at $t = 0$, so that optimal tax is zero.

Next consider the case where $b < c_*(0)$. Here community hostility induces too much compliance relative to the low social benefit of clean behavior. Once again, if $\beta'(t) = 0$, social loss is increasing in t, so that the optimal tax rate is zero. (Indeed, if negative rates of taxation were available – a subsidy for dirty behavior – they would be desirable to compensate the firm for the loss inflicted by an overly hostile population). However, if $\beta'(t) < 0$, we can construct numerical examples where social loss is decreasing in t in the neighborhood of $t = 0$, so positive rates of taxation may help to mitigate community hostility.

We summarize our findings for these two cases as:

Proposition 8 *(i) If community attitude does not depend on the tax rate (or is sufficiently insensitive to it) the optimal tax on dirty behavior is positive whenever $b > c^*(0)$ and non-positive whenever $b < c_*(0)$. (ii) These results may be reversed if community attitude is sufficiently sensitive to the tax rate.*

Finally we turn to the case where $c_*(0) < b < c^*(0)$, where

$$\begin{aligned}\frac{dSL(t)}{dt}_{|t=0,\ b\in(c_*,c^*)} = &-\tfrac{\gamma}{\beta+\gamma}(b-c_*)f(c_*)\tfrac{dc_*}{dt} + \tfrac{\beta}{\beta+\gamma}(c^*-b)f(c^*)\tfrac{dc^*}{dt} \\ &-\tfrac{\gamma\beta'(t)}{(\beta+\gamma)^2}\left[\int_{c_*}^{b}(b-c_i)f(c)dc - \int_{b}^{c^*}(c_i-b)f(c)dc\right].\end{aligned} \tag{28}$$

Here the welfare impact of taxation is ambiguous even when $\beta'(t) = 0$. The introduction of a tax encourages more clean behavior at the $c_*(0)$ margin, which decreases social loss (that is, increases welfare) since $c_* < b$. But it also encourages more clean behavior at the $c^*(0)$ margin, which lowers welfare because $b < c^*(0)$. The overall impact will depend upon the size of these two effects, and on the distribution $F(c)$: it cannot be determined in general. The ambiguity is further complicated by the possibility that β is itself sensitive to t.

Proposition 9 *If $c_*(0) < b < c^*(0)$ the imposition of a tax may increase or decrease welfare.*

Once again, a numerical example helps. Using the parameter configuration described earlier and assuming that $\rho = 0$ (so that β does not vary with t), social loss is increasing

in t when b is less than unity: in this case zero taxation is better than positive taxation. If, however, $b > 1$, social loss is decreasing in t, which calls for positive taxes.

6 Discussion

Optimistic commentators have promoted the idea that informal regulation – communities bringing pressure to bear upon firms – could replace more formal approaches in some settings, or usefully complement regulatory instruments such as taxation in others (Tietenberg (1998)). Such views underpin community right-to-know type policies.

Our results call into question the generality of these ideas. In our set-up the incentives generated by community pressure can never be as efficient as a well-functioning system of taxation, unless a particular coincidence of parameter values happens to pertain. Further those incentives may be welfare-reducing against a benchmark of no intervention. We also highlight the idea that complementarity between the formal instruments (in our case taxation) and informal pressure cannot be taken for granted – incentives may interact in a way that is unhelpful from a welfare perspective.

Further, in calibrating a tax (or other formal policy instrument) a policy-maker needs to make adjustment for the fact that community pressure may also be influencing pollution decisions. If, for example, a disclosure program is operating and effective, then the optimal emissions tax is not one necessarily calibrated to the standard Pigovian norm.

We adopted a Markov approach to modeling, though other approaches could have been used. The regime-switching view of community attitude – with the community potentially flip-flopping between support and hostility according to the recent (and perhaps not so recent) behavior of a firm – seems natural. The analysis incorporates two aspects of community pressure. First, that firm behavior may affect community attitudes in only a 'noisy' fashion. Whether an act of environmental transgression generates popular outrage may depends on public awareness of that act (what else was in the news that day?), on whether a critical mass of citizens are offended, and on the murky dynamics of 'group think'. Equally, whether or not community hostility is assuaged by subsequent environmental compliance may depend on practical details such as media management. This

inherent noisiness is well captured by a Markov process, with a firm's action determining the state of community attitude but only in a stochastic manner. Second, the penalty imposed by community hostility may be only loosely related to the true social cost of any environmental transgression. The channels through which a community may punish transgression (with-holding custom for a period, social ostracism of employees, etc.) are not subject to optimal calibration (or even measurement) in the way in which penalties in a formal regulatory setting could be. Of course formal regulation is not perfect and usually riddled by informational and other problems associated with costly enforcement. But at least in theory the penalties associated with formal regulation can be matched to firm choices.

There are limitations to the Markov approach. While community attitudes at any given moment are endogenous to the model, the process by which they evolve are guided by parameters – in our case, γ and β – that are not. They have not, for example, come out of any process of 'optimization' by the community. This seems like a reasonable starting assumption, given our limitations in understanding how community attitudes are shaped, but future work could focus on understanding these transition probabilities.

Our comparison of informal regulation with a tax regime was biased – the context was deliberately framed in such a way that tax could obtain first best – but instructive. The two-step pattern of clean behavior generated by community pressure was qualitatively distinct from the one-step pattern generated by the optimal tax and required for first-best. Of course in any real setting there may be other impediments to execution of an optimal tax regime, for example where regulatory governance is less than perfect. We can acknowledge that informal regulation may be particularly relied upon precisely in those sorts of settings.

How the incentives for clean behavior that community pressure generate may interact with other policy initiatives is an important thing to consider. There has been a presumption that when formal incentives are inadequate, informal incentives will simply 'top them up'. But this sort of additivity cannot be taken for granted – informal and formal interventions are not necessarily complementary. This theme would be worth exploring in the case of instruments other than the pollution taxation.

References

1. Badrinath, S. and P.J. Bolster (1996). "The role of market forces in EPA enforcement activity," *Journal of Regulatory Economics* 10(2): 165-81.

2. Baron, D.P. (2001). "Private politics, corporate social responsibility and integrated strategy," *Journal of Economics and Management Strategy* 10(1): 7-45.

3. Becker-Olson, K., A. Cudmore and R. Hill (2006). "The impact of perceived CSR on consumer behavior," *Journal of Business Research* 59(1): 46-53.

4. Biehl, A.R. (2001). "Durable-goods monopoly with stochastic values", *RAND Journal of Economics* 32(3): 565-77.

5. Blomberg, S. Brock, Gregory Hess and Akila Weerapana (2004). "Economic conditions and terrorism", *European Journal of Political Economy* 20(3): 463-78.

6. Brekke, K.A. and K. Nyborg (2008). "Attracting responsible employees: Green production as labor market screening", *Resource and Energy Economics* 39(7): 509-526.

7. Dasgupta, S., H. Hettige and D. Wheeler (2000). "What improves environmental compliance? Evidence from Mexican industry," *Journal of Environmental Economics and Management*, 39(1): 39-66.

8. Foulon, Jerome, Paul Lanoie and Benoit Laplante (2002), "Incentives for pollution control," *Journal of Environmental Economics & Management* 44(3): 169-87.

9. Gneezy, U. and A. Rustichini (2000), "A fine is a price," *Journal of Legal Studies* 29(1): 1-18.

10. Greenberg, Joseph (1984). "Avoiding tax avoidance," *Journal of Economic Theory* 32(1): 1-13.

11. Harrington, Winston (1988). "Enforcement leverage when penalties are restricted," *Journal of Public Economics* 37(1): 29-53.

12. Hettige, H., M. Huq, S. Pargal and D. Wheeler (1996), "Determinants of pollution abatement in developing countries: Evidence from South and Southeast Asia," *World Development* 24(12): 1891-1904.

13. Innes, R. (2006). "A theory of consumer boycotts," *Economic Journal* 116: 355-81.

14. Kassinis, G. and N. Vafeas (2006). "Stakeholder pressures for environmental performance," *Academy of Management Journal* 49(1): 145-59.

15. Klein, Jill, Craig Smith and Andrew John (2004). "Why we boycott: Consumer motivations for boycott participation," *Journal of Marketing* 68 (July): 92-109.

16. Kotchen, Matthew (2006). "Green markets and private provision of public goods," *Journal of Political Economy* 114(4): 816–834.

17. Landsberger, M. and I. Meilijson (1982). "Incentive generating state dependent penalty systems: The case of income tax evasion," *Journal of Public Economics* 19(3): 333-52.

18. Lagunoff, Roger (2006). "Credible communication in dynastic government", *Journal of Public Economics* 90(1): 59-86.

19. Lyon, T.P. and J.W. Maxwell (2008), "Corporate social responsibility and the environment: A Theoretical Perspective," *Review of Environmental Economics and Policy*, 2(2): 240-260.

20. Netzer, Oded (2008). "A hidden Markov model of customer relationship dynamics," *Marketing Science* 27: 185-204.

21. Pargal, Sheoli and David Wheeler (1996), "Informal regulation of industrial pollution in developing countries," *Journal of Political Economy* 104(6): 1314-1327.

22. Pargal, S., H. Hettige, M. Singh, and D. Wheeler (1997). "Formal and informal regulation of industrial pollution: Comparative evidence from Indonesia and the United States," *World Bank Economic Review* 11(5): 433-50.

23. Parsons, R. (2011). "BP ads fail to repair reputation", *Marketing Week*, May 2011.

24. Pearson, M. and P. West (2003). "Drifting smoke rings: social network analysis and Markov processes in a longitudinal study of friendship groups and risk taking," *Connections: bulletin of the International Network for Social Network Analysis* 25(2): 59-76.

25. Pelsmacker, P., L. Driesen and G. Rayp (2006). "Do consumers care about ethics: Willingness to pay for Fairtrade coffee," *Journal of Consumer Affairs* 39(2): 363-85.

26. Reinhardt, Forest L., Robert Stavins, and Richard Vietor (2008), "Corporate and social responsibility through an economic lens", *Review of Environmental Economics and Policy* 2(2): 219-239.

27. Roe B., M.F. Teisl, A. Levy, M. Russell (2001). "US consumers' willingness to pay for green electricity," *Energy Policy* 29(11):917-925.

28. Tietenberg, Tom (1998). "Disclosure strategies for pollution control," *Environmental and Resource Economics* 11(3): 587-602.

29. van Rooij, Benjamin (2010a). "Greening industry without enforcement? An assessment of the World Bank's pollution regulation model for developing countries," *Law & Policy* 32(1): 127-152.

30. van Rooij, Benjamin (2010b). "The People vs. Pollution: understanding citizen action against pollution in China", *Journal of Contemporary China* 19(63): 55-77.

31. Wang, Hua (2000). "Pollution charges, community pressure and abatement cost of industrial pollution in China," World Bank Policy Paper No. 2337.

32. Yu, Jingjhu and Jinli Pei (2009). "Study on the dynamic behavior of mass attitude," Bioinformatics 3rd International Conference Proceedings (ICBBE 2009).