

An Economic Model of Whistleblower Policy*

Anthony Heyes	Sandeep Kapur
Royal Holloway College	Birkbeck College
University of London	University of London

April 2007

Abstract

‘Whistleblowing’ is an increasingly common element of regulatory enforcement programs, and one that is encouraged by recent legislation in the US and elsewhere. We examine how responsive regulators should be to whistleblower tip-offs, and how severe should penalties be for wrongdoers detected in this way. Competing psychological theories as to what motivates employees to become whistleblowers are operationalized as alternative behavioral heuristics. Optimal policy depends upon the motives attributed to whistleblowers – which of the theories you subscribe to – but is not in general characterized by maximal penalties nor routine pursuit of complaints, even when pursuit is costless.

Keywords: Regulation, enforcement, behavioral law and economics

JEL Codes: K42

*We thank Thomas Ulen, Cees Withagen, Daniel Arce, Bernard Sinclair-Desagne, Charles Mason, Mark Raymond, Ron Smith, seminar participants at the LSE, Toulouse and the 2005 Industrial Organization Society Meetings (Atlanta), two anonymous referees and Editor Pablo Spiller for constructive comments. Errors are ours.

1 Introduction

The volume of ‘whistleblower’ activity in the United States and elsewhere has increased dramatically in recent years. Apart from the ‘celebrity’ cases that come readily to mind – Ingavar Bratt (who exposed plans for illegal weapons exports by the Swedish engineering firm Bofors), Cynthia Cooper and Sherrin Wadkins (who exposed accounting malpractice at WorldCom and Enron, respectively) and Chuck Hamel (who exposed lax safety practices and cut-backs in environmental precautions in the oil industry) – there has been a marked increase in the routine reporting of ongoing or prospective wrongdoing to regulators and enforcement agencies around the world.¹ The phenomenon straddles the public, private, and voluntary sectors, and almost every area of market and non-market activity.

The trend is expected to continue. For example, in the context of Medicaid fraud, in March 2007 the Attorney General’s Office in Louisiana was said to be anticipating an “explosion” in cases as a result of changes in policy designed to increase the number of whistleblowers (Lawley 2007).

The spurt in activity has been encouraged by the comparatively recent legislation designed to protect whistleblowers. The UK Public Interest Disclosure Act (1999) and the US Whistleblower Protection Act (1989) provide over-arching legal protection. The latter was strengthened by the Whistleblower Protection Enhancement Act (2007) approved by Congress in March 2007 and said by the US Government Accountability Project (GAP) to be “the most significant whistleblower rights legislation in US history.” Protections are also included in varying ways in a number of major pieces of regulatory legislation.² Section 806 of the Sarbanes-Oxley Act (2002) strengthened legal protection against retaliation for employees of public companies who report suspected corporate fraud. In 2004 the OECD agreed revised Principles of Corporate Governance, to include protection for whistleblowers (OECD 2004) and this has led to the reopening of policy debate in a number of OECD countries.

In 2001, the Public Service Commission of Canada published a comparative analysis of whistleblower legislation in various countries around the world, as a ‘primer’ to its reconsideration of whistleblower legislation (Groeneweg 2001). The general direction of movement is towards strengthening whistleblower protections and reducing the barriers that whistleblowers face when coming forward to pass incriminating information to enforcement agents.³

In general, the level of whistleblowing activity is likely to depend not just on the legal protection granted to whistleblowers but also on the regulatory response to whistleblowing. For instance, more prospective whistleblowers might come forward if the policy regime is deemed to be responsive – that is, if there is a higher probability that whistleblowing will trigger enforcement action rather than be ignored – or if the malfeasance uncovered by whistleblowing were expected to lead to severe penalties. This raises the question: What should an enforcement regime based on ‘tips’ from whistleblowers look like? In particular, how responsive should enforcement agencies be to information brought to them by whistleblowers, and how vigorous should be the punishment of malfeasance detected via this route?⁴

These two characteristics of the enforcement regime – responsiveness to whistleblowing and the size of penalties – matter not just because they influence the level of whistleblowing activity, but because they affect the ex-ante compliance decision. Despite the increased reliance put on whistleblowing in regulatory enforcement, there is no established economic model that incorporates it and so no framework within which to think systematically about regime design.⁵ This is a surprising absence, and one that we seek to correct.

In the context of an enforcement-compliance game – extended to incorporate whistleblowing, but with otherwise generic features – we explore the welfare-maximizing calibration of the enforcement regime. The pattern of whistleblowing activity is determined endogenously in the model and in setting policy the regulator has to assess how his choices impact the incentives facing both firms and prospective whistleblow-

ers. At first glance, one might be tempted to believe that the policy regime towards whistleblowing should be fully responsive and malfeasance should be penalized to the full extent the law permits. However, our model shows that this may not always be optimal, and that the correct choice of enforcement parameters depends crucially on the motives we ascribe to whistleblowers.

Glazer and Glazer (1989:4) define a whistleblower as one who (a) acts to prevent harm to others, not him or herself, (b) while possessing evidence that would convince a reasonable person. While other definitions are available, we adopt this frequently-cited one. For our purposes the critical aspect of whistleblowing is that there be scope for regulatory intervention to *prevent* harm or damage (which could include curtailing an ongoing pattern of misbehavior).⁶

The biggest hurdle in modeling whistleblowing – and perhaps a reason why the phenomenon has not proven amenable to economic analysis – follows directly from part (a) of the adopted definition. If the benefits from the activity accrue, by definition, to others and not the whistleblower, such behavior is not easily incorporated under conventional assumptions about rational, self-interested agents.⁷

We develop a behavioral model that adopts, in particular, the methods of behavioral law and economics. We construct a general whistleblower ‘motivation function’, which relates whistleblowers’ propensity to blow the whistle to the characteristics of the observed malfeasance and the enforcement environment. We provide evidence from sociology and psychology as to why employees blow the whistle on law-breaking employers, despite it not being in their narrowly-defined self-interest. We identify three broad (and quite distinct) schools of thought, and specialize the general motivation function to correspond to each school. We then characterize the optimal policy in each case. Optimal policy turns out to vary substantially between the cases.

Our analysis, then, is primarily normative in ambition. If you believe that school

X offers the most compelling explanation for whistleblower behavior – either in general, or in the specific setting of interest to you – then you should have a preference for policy design Y.

In Section 2 we describe a regulatory setting. We then provide a short summary of the three competing behavioral schools of thought, characterize and compare optimal enforcement policy under each. Section 3 relaxes the assumption that whistleblowers are fully knowledgeable: we assess the extent to which our conclusions regarding optimal enforcement policy are robust to the possibility of whistleblowers' errors. Section 4 draws out some of the policy implications, suggests how our analysis could inform future empirical work, and proposes ways in which our analysis could be extended. Section 5 concludes.

2 Model: A Regulatory Setting

We consider a single-shot compliance/enforcement setting in which whistleblowing can be embedded. Whilst the model has general applicability, for concreteness we write it up in the context of environmental malfeasance. Lieberman (2007:2) notes that

“Whistleblowers are particularly helpful in environmental settings. This is because many environmental violations and crimes are difficult to detect absent help from knowledgeable insiders. For example, if a company plans to lie about levels of some pollutants emitted into the air, often the only way that kind of lie can be detected is with the help of someone with inside information - a whistleblower.”

Recall that our definition of whistleblowing requires that the individual (a) acts to prevent harm to others, not him or herself, (b) while possessing evidence of intent that would convince a reasonable person.

To capture part (a) of the definition in a single-shot setting requires there be a lag between the decision to violate a law, and the moment at which the violation occurs and harm results. It is easy to think of stories consistent with this. An employee may be aware that a decision has been made to dispose of some waste illegally rather than using a licensed contractor. Before harm – in this case environmental damage – occurs the employee has the opportunity to report to the regulatory agency that violation is planned. Alternatively, in a repeated setting we might think of an ongoing behavior. It must also be the case that the regulator has time to respond to the report and prevent the plans for violation being executed (as in Heyes (1996)).

Consistent with (b), assume that the whistleblower has inside information regarding a firm’s intent to violate, information that an external agent could not observe directly but which can be evidenced. These are the un-shredded memos, e-mail exchanges and taped conversations of high-profile whistleblowing cases. Alternatively, the agent may not himself hold evidence, but be able to direct the investigator to the place in the organization where the evidence can be found. This is analogous to the assumption made by Prendergast (2003) in the context of consumer complaints regarding treatment by public bureaucracy. The regulator can summarily dismiss as spurious a complaint not supported by evidence.

Consider the following stylization. All firms are required to adhere to a regulatory standard. Non-compliance by a firm imposes external damage d , assumed to be the same across all firms. Compliance costs, on the other hand, differ across firms: firm i ’s cost of compliance, c_i , is drawn from a probability distribution on the interval $[0, \infty)$, with distribution function $G(c)$. Here $G(c)$ denotes the probability that a randomly-drawn firm has compliance cost c or lower. We assume that $G(d) < 1$, or that at least some firms have compliance costs that exceed d .

At time $t = 1$ each firm (which we can equally think of as a single principal) makes a plan either to ‘comply’ or to ‘violate’. That decision may be known to a number of employees, but to side-step complications due to multiple potential whistleblowers

we assume that the decision is known to a single agent. The agent holds evidence of the decision that is externally verifiable. To begin with, we assume that the agent also knows the firm's compliance cost c_i accurately, though information on costs is unverifiable. We consider the implications of relaxing this assumption of full knowledgeability in Section 3.

At $t = 2$ the agent has the opportunity to blow the whistle – that is, report the planned violation to a regulator.

At $t = 3$ the regulator decides whether to act upon any report received. With probability π the regulator ‘hears’ the report and sends an inspector to ‘visit’, and with probability $(1 - \pi)$ it ignores the report. Here π is a pre-determined and publicly-observed policy variable that captures the responsiveness of the regulator to whistleblowers. An enforcement visit entails two things. First, the firm is required to comply. Second, it is subject to a fine f for intent to violate. We assume that the fine cannot exceed some value F . Notice that whilst we will assume our regulatory agency is welfare-motivated it is restricted in the instruments it has available. For example, it cannot condone the violation in exchange for a fee. The two policy variables $\{\pi, f\}$ define the enforcement environment chosen by the regulator.

For simplicity we will proceed *as if* there are no pecuniary costs or benefits to the whistleblower. This assumption warrants some discussion. There are numerous legal protections for whistleblowers in the US and Europe, and the most restrictive assumption we could make would be that those protections worked perfectly. That, however, would not be realistic – there is plenty of evidence that legal protection is less than full and that whistleblowing imposes a significant costs on individuals. Nonetheless, the key costs upon the individual are usually redistributive in nature – the loss of the nicer task, promotion, preferred office, salary increase *etc.*. The social costs appear slight. In terms of behavior, then, the private costs can be regarded as captured in the $\mu_i, \theta_i, \delta_i$ ‘thresholds’ that we will introduce in Sections 2.3 through 2.5. In terms of welfare analysis they are transfers and can be ignored.

The recently-completed case relating to M/V Irika, a 623-foot ocean-going bulk carrier, is an example of the sort of situation captured in our model:

On October 5, 2006, U.S. Coast Guard inspectors from Coast Guard Sector Portland boarded the ship at Vancouver, Washington to conduct a routine inspection. Inspectors reviewed the ship's Oil Record Book, a document in which ship engineers must record all transfers and discharges of oil. Inspectors were unable to identify any discrepancies but were subsequently contacted by the ship's Second Engineer who discretely told the inspectors about a flexible hose being used to bypass pollution prevention equipment oily sludge directly into the ocean. This prompted inspectors to conduct a second inspection on October 6, during which they located the hose and found additional evidence of its use. As a result of the prosecution the owners (Irika Maritime) were fined \$750,000 and placed on four years' probation. While on probation they must implement an Environmental Management System/Compliance Plan (EMS) to ensure there is no future dumping. The judge noted that the whistleblower should be protected from reprisal, but held no expectation of financial reward when contacting the inspectors. (US Attorney's Office 2007: 2).

We capture the whistleblowers' propensity to disclose in the form of a motivation function. In the most general case we allow the disclosure decision to depend on compliance cost c_i and the enforcement parameters π and f , the three 'moving parts' in our model. The motivation function $\rho(c_i, \pi, f)$ measures the probability that an individual drawn at random from the population of prospective whistleblowers will disclose when faced with the particular configuration of these variables.

2.1 Policy Problem

We assume that the regulator acts to minimize social loss, defined as the sum of expected compliance costs and external damage.⁸ His problem, then, is to choose $0 \leq \pi \leq 1$ and $f \leq F$ to minimize:

$$SL(\pi, f) = \int_{Comply} c_i dG(c) + \int_{Violate} [\rho(c_i, \pi, f)\pi c_i + (1 - \rho(c_i, \pi, f)\pi)d] dG(c) \quad (1)$$

where firm i violates if and only if its cost of compliance exceeds the expected penalty associated with non-compliance:

$$c_i > \rho(c_i, \pi, f)\pi(c_i + f). \quad (2)$$

The first integral in (1) aggregates compliance costs across the set of firms that makes an initial decision to comply. The second integral aggregates compliance costs and external damage across firms that choose initially to violate: the actual loss in the latter case depends on whistleblower activity and regulatory responsiveness in the natural way.

Thus the policy question is two-fold: (i) How responsive should the regulatory agency be to reports from whistleblowers? (ii) how severely should firms caught through such disclosures be punished? In other words, to choose π and f .

Analysis of this general problem allows remarks to be made with respect to families of functions but at the cost of tractability. Here we choose to work with three alternative theories regarding whistleblower motivation, each associated with one of the ‘schools of thought’ identified below (and we will use the subscript ρ_Δ , $\Delta \in \{\alpha, \beta, \gamma\}$, to denote the whistleblower motivation function associated with the theory ‘in play’).

2.2 Whistleblower Motivation: Three Theories

What does research in other disciplines say about the motivations that might underpin the behavior of whistleblowers?

The first (and preponderant) school of thought is that whistleblowing constitutes **conscience cleansing**. The hypothesis is that individuals have moral codes that condition their behavior. In our context, the moral code of an agent (employee) may preclude him from being complicit with, ‘going along with’ or ‘keeping quiet about’ immoral activities proposed by his principal (employer). Being complicit with an immoral activity corrupts the self and “... (w)histleblowers disclose because they dread living with a corrupted self more than they dread the other outcomes” (Alford (2001: 90)).⁹ The agent will not comply unquestioningly with an instruction issued by the principal but will, rather, subject that instruction to a moral screen. Authors subscribing to and providing evidence for this view include Miceli and Near (1992) and Jos et al. (1989) among others.

As we discuss later, the determination of the ‘morality’ or otherwise of a particular activity may depend on cultural, religious and other forces. Importantly, the decision by an employee to disclose any planned illegal activity of an employer depends only upon the *characteristics of the planned activity itself*. If an activity is sufficiently ‘bad’ the agent is morally unable from being complicit with and feels compelled to speak out to cleanse his conscience. Significantly for policy design, in this case the whistleblowing decision is *independent of the characteristics of the enforcement regime*. Variants of this theory are popular amongst social psychologists and sociologists and are consistent with the voluminous case-study evidence that people engage in whistleblowing even where they know that the likelihood that they will be ‘heard’ – that their blowing the whistle will impact outcomes – is minimal.¹⁰

In order to operationalize this in our model we need some measure of the moral defensibility of the act of violation. The prospective violator, recall, chooses to impose external environmental damage of value d in order to save private costs c . The relationship between c and d embodies the rate at which the employer trades-off the interests of self versus others in his decision. Since the impact of non-compliance is fixed at d , we can simply regard c_i as a measure of the defensibility of the firm’s

decision to violate. If c were a very large number, then it would require very large self-sacrifice on the part of the employer to avoid imposing the damage. A reasonable person might take a view that non-compliance with the regulation is more defensible when c is large than when it is a small number. We might then assume that each individual has some threshold of defensibility beyond which his conscience compels him to speak out.¹¹ We operationalize this as

Assumption α : Individual i reports planned violation if and only if $c_i < \mu_i$.

It is plausible that the threshold μ will vary privately across individuals. We introduce a distribution function for this threshold in the analysis that follows. Importantly, in this case the regulator will not be able to change the pattern of whistleblowers coming forward by changing the enforcement regime.

As an alternative hypothesis, individual moral motivation could have a **welfarist** basis such that the whistleblowing decision will incorporate a cost-benefit element. Hunt (1997) proposes that a ‘justifiable disclosure’ must at minimum (i) serve some purpose in correcting or preventing harm, and (ii) do more harm than good. Our chosen definition of whistleblowing implies that (i) is necessarily satisfied, whilst (ii) implies both that the whistleblowing decision is forward-looking and that the agent applies some sort of cost-benefit test. The welfarist motivation is also consistent with the popular economic characterization of altruism as ‘other-regarding behavior’ – that the utility of others may enter as an argument in the agents’ own utility function – and the closely related notion of ‘warm glow’ (Andreoni 1990).

This second behavioral theory, then, proposes a utilitarian or social welfare basis for disclosure. We operationalize this by assuming that an individual will disclose if the expected *social* benefits from so doing are sufficiently large.

Assumption β : Individual i discloses planned violation if and only if $\pi(d - c_i) > \theta_i$.

Here $\theta_i \geq 0$ is an individual-specific threshold that captures what individual i regards as ‘sufficient’.¹² Under this assumption agents are more pragmatic, and

look to the likely impact of their actions. A potential whistleblower takes the responsiveness parameter π into account when making his decision (since that affects the welfare return) but does not take account of the penalty f for non-compliance (which is a simple transfer). Thus the disclosure decision is itself sensitive to (one dimension of) the enforcement regime.

A third view – and one popular in the ‘airport’ literature – is that whistleblowers are **punishment motivated**, or what Alford calls ‘hysterical malcontents’, primarily driven by a desire to discomfort the organization that employs them.

A certain proportion of staff might simply be ‘disgruntled employees’, unhappy for reasons unconnected with the firms planned non-compliance with the regulation, but opportunistic in blowing the whistle when so doing creates sufficiently substantial discomfort (cost) for their employer. Alternatively, the agent’s desire to punish his principal may relate to the non-compliance decision itself. There is compelling experimental evidence that a portion of subjects are willing to act to punish others who have behaved anti-socially or illegally, even if that punishment is not in their apparent self-interest nor the interest of *prospective* social welfare (Fehr and Schmidt (1999), Fehr and Falk (2002)).

We operationalize this notion of punishment-motivated disclosure by proposing that the agent will disclose only if the damage he can do to the firm is sufficiently large.

Assumption γ : An employee discloses planned violation if and only if the expected cost impact upon the firm is sufficiently large: $\pi(c_i + f) > \delta_i$.

Once again, δ_i is an individual-specific threshold. Observe here that both π and f impact upon the whistleblowers’ decision.

2.3 Policy if Whistleblowing is Conscience Cleansing

Under Assumption α individual i reports planned violation if and only if $c_i < \mu_i$. We allow the threshold μ to vary privately across individuals and denote the probability distribution of this attribute in the population as $\Phi_\mu(\cdot)$. If so, the probability that planned violation by firm i will be disclosed to the regulator depends functionally on the defensibility of the act:

$$\rho_\alpha(c_i) = 1 - \Phi_\mu(c_i).$$

It is important here – recalling the discussion of conscience cleansing above – to note that $\{\pi, f\}$, the characteristics of enforcement environment, do not impact the disclosure decision. The whistleblower takes a view on the morality of his employer’s choice, and decides whether or not his conscience allows him to ‘live with’ or be complicit with that choice, without reference to the enforcement environment.

< Figure 1 here >

In Figure 1 we represent $\rho_\alpha(c_i)$ as an ogive, consistent with μ being drawn from a continuous, single-peaked distribution, but nothing rests on this.

In this environment, firm i will comply voluntarily if

$$c_i \leq \rho_\alpha(c_i)\pi(c_i + f). \quad (3)$$

Compliance incentives – the net expected benefits of compliance – decrease monotonically in c_i . That is, $[\rho_\alpha(c_i)\pi(c_i + f) - c_i]$ is decreasing in c_i . An increase in c_i both makes compliance more costly and reduces the chance that non-compliance will be reported and penalized. For values of c_i close to zero net benefits must be positive. Therefore, the firm will comply if and only if its compliance cost is less than some critical value $\widehat{c}(\pi, f|\alpha)$, implicitly defined by

$$\rho_\alpha(\widehat{c})\pi(\widehat{c} + f) - \widehat{c} = 0. \quad (4)$$

It is straight-forward to confirm that \widehat{c} is increasing both in π and in f . Firms with compliance costs less than \widehat{c} comply voluntarily. Firms with costs above \widehat{c} comply only if coerced, which occurs with probability $\rho_\alpha(c_i)\pi$. In summary:

Remark 1 *Under Assumption α the net private benefits from compliance decrease monotonically in the cost of compliance. For a given enforcement environment $\{\pi, f\}$ the realized probability of compliance is non-increasing in c_i . It equals 1 if $c_i \leq \widehat{c}$, and equals $\rho_\alpha(c_i)\pi$ otherwise.*

Non-compliant firms impose external damage d . The regulator's problem is to choose $0 \leq \pi \leq 1$ and $f \leq F$ to minimize social loss:

$$SL(\pi, f|\alpha) = \int_0^{\widehat{c}} c_i dG(c) + \int_{\widehat{c}}^\infty [\rho_\alpha(c_i)\pi c_i + (1 - \rho_\alpha(c_i)\pi)d] dG(c). \quad (5)$$

We begin with the observation that a welfare-motivated regulatory agency will never choose to implement a regime that induces $\widehat{c} \geq d$. Such a regime induces compliance beyond that which is justified by external damage d and is welfare-reducing.¹³ Given this, consider any pair $\{\pi', f'\}$ such that $\widehat{c} < d$. Then,

$$\frac{\partial SL(\pi', f'|\alpha)}{\partial f} = \frac{\partial \widehat{c}}{\partial f} (1 - \rho_\alpha(\widehat{c})\pi)(\widehat{c} - d) < 0. \quad (6)$$

A regulator that aims to minimize social loss must then choose the maximal penalty available to it, or that optimal policy entails $f^* = F$.

An interior solution π^* to the regulator's choice of the responsiveness parameter will then be characterized by

$$\frac{\partial SL(\pi^*, F|\alpha)}{\partial \pi} = 0 = \frac{\partial \widehat{c}}{\partial \pi} (1 - \rho_\alpha(\widehat{c})\pi^*)(\widehat{c} - d) + \int_{\widehat{c}}^\infty \rho_\alpha(c_i)(c_i - d) dG(c), \quad (7)$$

where $\widehat{c} = \widehat{c}(\pi^*, F|\alpha)$. The first composite term on the right-hand side of (7) is negative: An increase in inspection probability extends the interval of firms that plan to comply, which at the $\widehat{c} < d$ margin increases welfare. The second term captures the social loss associated with the increased frequency with which compliance will

be coerced at firms with compliance costs above \hat{c} , which must be positive in the vicinity of an interior solution.

If the expression in (7) is negative when evaluated at $\pi = 1$ then the corner solution $\pi^* = 1$ results. For current purposes we will assume that this upper corner solution is not binding, which amounts to assuming that the instruments of enforcement are sufficiently potent.

Proposition 1 *Assume that (a) whistleblowing is motivated by conscience cleansing and (b) the regulator's enforcement instruments are potent enough that $\hat{c}(1, F|\alpha) > d$. Then optimal policy will be characterized by (i) a maximal penalty and (ii) an inspection intensity set less than maximally. This applies even though inspection is costless.*

A well-known result in the economics of enforcement, attributed to Becker (1968), argues that to achieve any particular level of compliance the agency should set fines maximally and adjust inspection intensity correspondingly (see Heyes (2000) for a survey of the literature that has developed around the Becker model). Becker's result is based on inspections being costly, while fines are not. A similar result holds here, where inspections are costless, but for a different reason. In raising *voluntary* compliance to any particular level there is a strict preference based not on consideration of enforcement costs, but on the qualitative pattern of *actual* compliance that results.

In our model the regulator can increase compliance by raising either its responsiveness π to whistleblowers' or by raising the penalty f for non-compliance. Doing so through increases in f will always be beneficial provided \hat{c} remains below d . An increase in f has an effect only at the margin, and the additional voluntary compliance induced at that margin is welfare-improving. Raising \hat{c} through increases in π is different. Again, at the margin the additional voluntary compliance is beneficial provided \hat{c} remains below d . But there is a welfare cost of so doing, which is the

increased frequency with which compliance is coerced at firms where compliance is welfare-reducing. Equation (7) specifies that, with f fixed at its maximal value F , the agency sets π to trade-off the net benefits from increased voluntary compliance at the margin with the net disbenefits from increased coerced compliance.

Given that the pursuit of reports is costless in our model, the choice of π and f depends only upon the efficiency of the pattern of compliance that is realized. First-best would be described by compliance with probability 1 for $c_i \leq d$, zero otherwise. Against this benchmark we note that in equilibrium firms can be divided into three classes. Those with compliance costs $c_i \leq \widehat{c}(\pi^*, F|\alpha)$ will comply pre-emptively, and that compliance will be socially desirable. The mid-range interval of firms $\widehat{c}(\pi^*, F|\alpha) < c_i \leq d$ comply only when coerced (that is, with probability $\rho_\alpha(c_i)\pi$), and that compliance is socially desirable. The interval of high-cost firms $c_i > d$ also comply only when coerced, but that compliance is socially undesirable. The comparison between realized compliance probabilities given optimal policy (solid line) and the first-best pattern (broken line) is illustrated in Figure 2. In summary,

< Figure 2 here >

Remark 2 *Under Assumption α optimal policy implements an equilibrium in which an interval of low cost firms comply and this is efficient. An interval of high cost firms comply too often, an interval of intermediate cost firms comply not often enough.*

2.4 Policy if Whistleblowing is a Social Act

Under Assumption β individual i discloses planned violation if and only if

$$\pi(d - c_i) > \theta_i.$$

Let $\Phi_\theta(\cdot)$ be the distribution of the attribute θ in the population. Then the probability that a planned violation by firm i will be disclosed is given by $\Phi_\theta(\pi(d - c))$.

We denote this as $\rho_\beta(c_i, \pi)$. We might plausibly think of $\rho_\beta(c_i, \pi)$ as an ogive in c_i – with θ_i distributed according to some single-peaked distribution – but again nothing rests on this.

Given that $\theta_i \geq 0$, we have $\rho_\beta(c_i, \pi) = 0$ for $c_i \geq d$: whistleblowers motivated by social welfare would not disclose planned violations where compliance costs exceed the external damage. More generally, $\partial\rho_\beta(c_i, \pi)/\partial c_i \leq 0$ so that firms with higher compliance costs face lower risk of disclosure. Further $\partial\rho_\beta(c_i, \pi)/\partial\pi \geq 0$: other things equal an increase in the regulator’s responsiveness to disclosures increases the probability of disclosures.

Firm i will comply voluntarily if

$$c_i \leq \rho_\beta(c_i, \pi)\pi(c_i + f) \quad (8)$$

As under the previous assumption, it is straightforward to establish that the net expected benefits from compliance are everywhere (weakly) decreasing in c_i . Firm i complies voluntarily if c_i is less than some critical value $\hat{c}(\pi, f|\beta) < d$. Once again, \hat{c} is increasing in π and in f . Everything again works in the ‘right’ direction, allowing us to replicate Remark 1.

Remark 3 *Under Assumption β the net private benefits from compliance decrease monotonically in the cost of compliance. For a given enforcement environment $\{\pi, f\}$ the realized probability of compliance is non-increasing in c_i : it equals 1 if $c_i \leq \hat{c}$, and equals $\rho_\beta(c_i, \pi)\pi$ otherwise.*

The regulator sets policy to minimize expected social loss subject to the behavioral Assumption β , and the self-selection behavior of firms. That is, it chooses $0 \leq \pi \leq 1$ and $f \leq F$ to minimize

$$SL(\pi, f|\beta) = \int_0^{\hat{c}} c_i dG(c) + \int_{\hat{c}}^\infty [\rho_\beta(c_i, \pi)\pi c_i + (1 - \rho_\beta(c_i, \pi))d] dG(c) \quad (9)$$

For this case, $\rho_\beta = 0$ for $c_i > d$: a complaint from a whistleblower arises only arise when the cost of compliance is less than the damage avoided. As ρ_β is everywhere

non-decreasing in π and invariant to f and as $\partial\hat{c}/\partial\pi$ and $\partial\hat{c}/f$ are positive, it is clear that the agency can do no better than set $\pi = 1$ and $f = F$.

Proposition 2 *Assume that whistleblowing is motivated by social welfare. Then optimal policy will be characterized by (i) a maximal penalty and (ii) a maximal inspection intensity.*

The logic is straight-forward. Whistleblowers in this scenario have the same interests as the regulator – prospective social welfare. Whenever a report is forthcoming it will relate to a case in which $c_i < d$ so the regulatory agency should pursue it ($\pi^* = 1$). No firm with a compliance cost greater than d will ever be induced to comply because it knows it will never be reported. The regulator will wish to bring \hat{c} as close to d as he can, implying that he will set f as high as he can. Naturally this result might be altered if inspections are costly.

In terms of the resulting pattern of compliance,

Remark 4 *Under Assumption β optimal policy implements an equilibrium in which an interval of low cost firms comply and this is efficient. An interval of high cost firms never complies, and this is efficient. If $\hat{c}(1, F|\beta) < d$ an interval of intermediate cost firms comply not often enough.*

2.5 Policy if Whistleblowing is by Disgruntled Employees

Under Assumption γ an employee discloses planned violation if and only if the expected cost impact upon the firm is sufficiently large:

$$\pi(c_i + f) > \delta_i.$$

Again, the threshold δ_i can be interpreted as varying across individuals, with distribution function $\Phi_\delta(\cdot)$. The probability that a planned violation will be disclosed equals $\Phi_\delta(\pi(c_i + f))$. We denote this as $\rho_\gamma(c_i, \pi, f)$.

An immediate implication of this assumption is that the probability that a planned violation will be disclosed is *increasing* in the cost of compliance, or that $\partial\rho_\gamma/\partial c_i \geq 0$. It is also now increasing in both enforcement parameters: $\partial\rho_\gamma/\partial\pi \geq 0$ and $\partial\rho_\gamma/\partial f \geq 0$. Prospective whistleblowers want to punish, and so increased expected penalty increases the likelihood of report. In contrast to earlier cases, f now matters not just for the firm's compliance incentives directly, but also indirectly through its impact on the reporting behavior of whistleblowers.

Given enforcement environment $\{\pi, f\}$, firm i 's net benefits to compliance are:

$$\rho_\gamma(c_i, \pi, f)\pi(c_i + f) - c_i \quad (10)$$

Recall that under Assumptions α and β , net benefits to compliance were monotonically decreasing in c – an increase in c made pre-emptive compliance more costly, and also made non-compliance less likely to be reported. However, here differentiating (10) with respect to c gives:

$$\pi \left[\frac{\partial\rho_\gamma(c_i, \pi, f)}{\partial c_i}(c_i + f) + \rho_\gamma(c_i, \pi, f) \right] - 1 \stackrel{\geq}{\leq} 0. \quad (11)$$

An increase in c makes pre-emptive compliance more costly, but now *increases* the likelihood that any given decision not to comply will be reported. The impact on net benefit is ambiguous.

Remark 5 *Under Assumption γ the net private benefits from compliance may increase or decrease with cost of compliance.*

In further contrast to the earlier cases:

Proposition 3 *Realized non-compliance probabilities may be non-monotonic in cost of compliance.*

We prove this by means of a simple degenerate example in which $\delta_i = \delta$ for all i . In that case $\rho_\gamma(c_i, \pi, f) = 1$ if $c_i > \left[\frac{\delta}{\pi} - f\right]$, equals zero otherwise. Assume that

parameters are such that $\frac{\pi}{(1-\pi)}f > \frac{\delta}{\pi} - f$. In that case firms comply voluntarily if and only if

$$\frac{\pi}{(1-\pi)}f > c_i > \frac{\delta}{\pi} - f$$

For c_i outside this mid-range the firm will only comply if coerced, which will happen with probability π for $c_i > \frac{\pi}{(1-\pi)}f$, and with probability zero for $c_i < \frac{\delta}{\pi} - f$ (Figure 3). In this degenerate case the firms above the mid-range do not comply pre-emptively because so doing is too expensive and it is more attractive to wait to see if they will be coerced. Firms below the range do not comply pre-emptively because their low costs places them below the point at which their non-compliance will be disclosed. In a non-degenerate version the low propensity to comply at the lower end would be driven by the comparatively low likelihood of non-compliance being reported.

< Figure 3 here >

If we restrict attention to settings in which the net benefits from compliance are monotonically decreasing in c_i (that is, where the additional up-front cost effect outweighs the increased probability of report effect) then firm i will pre-emptively comply if c_i is less than $\hat{c}(\pi, f|\gamma)$, implicitly defined by

$$\hat{c} = \rho_\gamma(\hat{c}, \pi, f)\pi(\hat{c} + f). \quad (12)$$

An argument similar to that used earlier can establish that optimal policy will involve implementing some $\hat{c} < d$. We can no longer argue, however, that for any combination of π' and f' (such as the optimal one) generating $\hat{c} < d$ that social loss will be decreasing in f . To see this note that

$$\frac{\partial SL(\pi', f'|\gamma)}{\partial f} = \frac{\partial \hat{c}}{\partial f} \cdot (1 - \pi \rho_\gamma(\hat{c}, \pi, f'))(\hat{c} - d) + \int_{\hat{c}}^{\infty} \pi \frac{\partial \rho_\gamma(\hat{c}, \pi, f')}{\partial f} (c - d) \quad (13)$$

where $\hat{c} = \hat{c}(\pi', f'|\gamma) < d$. The first term, under Assumption γ , is negative. The second term is ambiguous in sign and will be positive in the vicinity of an interior

solution to the regulator’s problem. Without further restriction on ρ this cannot be signed. This leads naturally to the following:

Proposition 4 *If whistleblowing is punishment motivated, optimal policy may involve (i) a less-than-maximal penalty and/or (ii) a less-than-maximal inspection intensity.*

The unusual element here is that f may be less than maximal. Under Assumptions α and β , having established that equilibrium would involve $\hat{c} < d$, it was simple to understand why the instrument f should be raised to its maximal level. Increases in f increased voluntary compliance at the \hat{c} (welfare-enhancing) without any cost. Under γ , however, increases in f now impact compliance away from the margin, increasing the frequency with which planned violations at high-cost firms are reported and compliance coerced. This effect may lead the regulator to wish to refrain from setting f at the highest level.

3 Whistleblowers with Noisy Information

Our analysis thus far assumes that whistleblowers are knowledgeable. This ignores the very real possibility that whistleblowers base their disclosure decision on faulty or partial information about the underlying firm-specific parameters. In terms of the moving parts of our model, it could be that whistleblowers under- or over-estimate the compliance cost of the firm. That way, for example, a conscience-cleansing employee might mis-evaluate the egregiousness of his employer’s actions, and make his report/not report decision on the basis of that mis-evaluation.

We develop this possibility by assuming that the potential whistleblower’s observation of compliance cost is noisy: a whistle-blower observes

$$\tilde{c}_i = c_i + \varepsilon_i, \tag{14}$$

where ε denotes noise. Negative values of ε reflect situations in which the whistleblower underestimates the true compliance cost while positive values reflect overestimation. For tractability, we assume that the distribution of the noise term, $E(\varepsilon)$ is common knowledge and is independent of the firm's true cost c_i . We also assume that the distribution of noise is independent of the prospective whistleblowers' characteristics, as captured by the distributions Φ_μ , Φ_θ and Φ_δ .

Whistleblowers' disclosure decisions now depend on the noisily-observed value of compliance costs. Errors in observation that lead them to over- or under-estimate costs, depending on the scenario, alter the probability of disclosure. The likelihood of disclosure depends on the realization of the error term, so that in general we now have $\rho(c_i, \pi, f|\varepsilon_i)$.

The firm does not observe the realization of the error term but knows its distribution. Given compliance cost c_i , it evaluates the likelihood of disclosure as

$$\rho(c_i, \pi, f) = \int \rho(c_i, \pi, f|\varepsilon)dE(\varepsilon). \quad (15)$$

As we argue below, noisy observation of compliance cost may raise or lower the probability of disclosure relative to the case without noise. As before, a firm's incentive to comply pre-emptively depends on its estimated risk of disclosure. It will comply pre-emptively if and only if

$$c_i \leq \rho(c_i, \pi, f)\pi(c_i + f). \quad (16)$$

Social welfare varies with actual reports of violations, so depends directly on the realization of ε . For any realization of ε , define

$$SL(\varepsilon) = \int_{Comply} c_i dG(c) + \int_{Violate} [\rho(c_i + \varepsilon)\pi c + (1 - \rho(c_i + \varepsilon)\pi)d]dG(c). \quad (17)$$

As before, the regulator's problem is to choose π and f to minimize expected social loss (that is, its expectation across possible realizations of the error term):

$$SL(\pi, f) = \int SL(\varepsilon)dE(\varepsilon). \quad (18)$$

We adapt this general framework to each of the environments studied earlier, allowing us to examine how noise in whistleblowers' information affects optimal enforcement policy.

3.1 Assumption α

Under Assumption α , an agent with noisy information reports a planned violation if and only if $\tilde{c}_i < \mu_i$. Given distribution Φ_μ of characteristic μ , the probability of disclosure conditional on noise ε is

$$\rho_\alpha(c_i|\varepsilon_i) = 1 - \Phi_\mu(c_i + \varepsilon_i). \quad (19)$$

For this case a potential whistleblower who underestimates compliance cost is more likely to report a violation than one who overestimates it. The firm computes the likelihood of disclosure by taking expectation over error terms:

$$\rho_\alpha(c_i) = 1 - \int \Phi_\mu(c_i + \varepsilon) dE(\varepsilon). \quad (20)$$

In general, relative to the case without noise, the noisy observation of compliance cost may raise or lower the perceived likelihood of disclosure. Clearly, if the agent always underestimates compliance costs (that is, if the support of distribution E is negative), the expected probability of disclosure is unambiguously greater. On the other hand, systematic overestimation of compliance costs lower the risk of disclosure. The case where errors are unbiased – consider a distribution of error terms with zero mean and positive variance – has ambiguous effects: If $(c + \varepsilon)$ is a mean-preserving spread of c , it follows that $\int \Phi_\mu(c_i + \varepsilon_i) dE(\varepsilon)$ is less than $\Phi_\mu(c_i)$ if Φ_μ is locally concave. In that case, noise will cause expected probability of disclosure to rise. For convex ranges of the distribution Φ_μ , noise will cause expected likelihood of disclosure to fall.

As in the case without noise, net expected benefit from compliance is decreasing in c_i so that firm i will comply voluntarily if and only if c_i is less than some critical

value $\widehat{c}(\pi, f)$. This critical value depends on the distribution of noise terms but does not depend on the actual realization of ε . As before it is straightforward to confirm that $\partial\widehat{c}/\partial\pi$ and $\partial\widehat{c}/\partial f$ are both positive.

Social loss depends on reported violations, so varies with the realization of ε . The regulator must choose enforcement parameters $\{\pi, f\}$ to minimize $SL(\pi, f|\alpha) = \int SL(\varepsilon|\alpha)dE(\varepsilon)$ where

$$SL(\varepsilon|\alpha) = \int_0^{\widehat{c}} cdG(c) + \int_{\widehat{c}}^{\infty} [\rho_{\alpha}(c + \varepsilon)\pi c + (1 - \rho_{\alpha}(c + \varepsilon)\pi)d]dG(c). \quad (21)$$

Replicating the familiar argument we can show that the welfare-motivated agency will never implement a regime that induces $\widehat{c} \geq d$. Then, for any pair of enforcement parameters that satisfy the restriction $\widehat{c} < d$, the partial derivative

$$\frac{\partial SL(\cdot|\alpha)}{\partial f} = \int \frac{\partial\widehat{c}}{\partial f} [(1 - \rho_{\alpha}(\widehat{c}|\varepsilon, \cdot)\pi)(\widehat{c} - d)]dE(\varepsilon)$$

is negative, so that the corner solution with the maximal penalty $f^* = F$ is optimal.

An interior solution to the agency's choice of π is given by

$$\frac{\partial SL(\pi^*, F|\alpha)}{\partial \pi} = 0,$$

which now requires

$$\frac{\partial\widehat{c}}{\partial \pi} [(1 - \rho_{\alpha}(\widehat{c})\pi^*)(\widehat{c} - d)] + \int \int_{\widehat{c}}^{\infty} \rho_{\alpha}(c|\varepsilon)(c - d)dG(c)dE(\varepsilon) = 0. \quad (22)$$

As in the case without noise, the first composite term in the expression is negative and the second must be positive at any interior solution for π . The logic of Proposition 1 – that if whistleblowing is motivated by conscience cleansing and if enforcement instruments are sufficiently potent – optimal policy involves maximal penalty but less than maximal responsiveness to whistleblowing – carries over to the case where whistleblowing is based on noisy information.

3.2 Assumption β

Under Assumption β , an agent with a noisy observation $\tilde{c}_i = c_i + \varepsilon$ reports a planned violation if and only if $\pi(d - \tilde{c}_i) > \theta_i$. Given distribution of characteristics Φ_{θ} , the

probability of disclosure conditional on ε is

$$\rho_\beta(c_i, \pi | \varepsilon_i) = \Phi_\theta(\pi(d - (c_i + \varepsilon_i))). \quad (23)$$

The firm evaluates the expected likelihood of disclosure as

$$\rho_\beta(c_i, \pi) = \int \Phi_\theta(\pi(d - (c_i + \varepsilon))) dE(\varepsilon). \quad (24)$$

As under Assumption α errors in observing compliance costs may increase or decrease the expected likelihood of being reported relative to the case without noise, depending on the distribution of error terms and the local concavity /convexity of the distribution function Φ_θ . Also, as under Assumption β in the absence of noise, the net benefit of compliance is decreasing in c so that voluntary compliance is rational for all c less than some critical value $\hat{c}(\pi, f|\beta)$. This threshold value depends on the distribution of noise terms and not on the actual realization of ε . It is straightforward to confirm that \hat{c} is increasing in π and in f .

Social loss depends on actual disclosures, so varies with ε . Define

$$SL(\varepsilon|\beta) = \int_0^{\hat{c}} c dG(c) + \int_{\hat{c}}^\infty [\rho_\beta(c + \varepsilon, \pi)\pi c + (1 - \rho_\beta(c + \varepsilon, \pi)\pi)d] dG(c).$$

The regulator minimizes the expected social loss, $SL(\pi, f|\beta) = \int SL(\varepsilon|\beta) dE(\varepsilon)$. Revisiting arguments established earlier under Assumption β in the absence of noise, we can confirm that (i) optimally-chosen enforcement parameters must implement a regime that has $\hat{c} < d$; (ii) the maximal penalty is still optimal; (iii) however, maximal responsiveness to whistleblowing may no longer be optimal.

To appreciate the last part, note that the regulator realizes that it is no longer the case – as was in the absence of noise – that complaints arise only for $c_i < d$. Specifically, there is a range of possibilities with $c_i > d$ where enforcement does not increase social welfare, yet a complaint based on underestimated compliance costs (a negative realization of ε such that $c_i + \varepsilon < d$) coerces compliance. In setting the optimal enforcement parameters, the regulator must take account of this possibility, possibly by lowering the inspection probability π below its maximal value.¹⁴ This is

hardly surprising. Consider the case where the variance of the noise term is so large that the informational content of whistleblowing is low: if so, it cannot be optimal for a welfare-motivated regulator to respond to every disclosure.

3.3 Assumption γ

Under Assumption γ , an agent with a noisy observation $\tilde{c}_i = c_i + \varepsilon$ reports a planned violation if and only if $\pi(\tilde{c}_i + f) > \delta_i$. Conditional on the noise term ε , the probability of disclosure now equals

$$\rho_\gamma(c_i, \pi, f | \varepsilon_i) = \Phi_\gamma(\pi(c_i + \varepsilon_i + f)). \quad (25)$$

For any level of compliance cost c_i , the firm evaluates the expected likelihood of disclosure as

$$\rho_\gamma(c_i, \pi, f) = \int \Phi_\gamma(\pi(c_i + \varepsilon + f)) dE(\varepsilon). \quad (26)$$

The introduction of noise in estimates of compliance costs does not alter the qualitative findings relative to the case without noise. As we found in Section 2.5, the net benefit from compliance may not be monotonic in c . Restricting attention to cases in which the net benefit is decreasing in compliance costs, there will be voluntary compliance for c_i below some threshold $\tilde{c}(\pi, f | \gamma)$. Assuming that the regulator chooses $\{\pi, f\}$ to minimize expected social loss, optimal policy may involve a less than maximal penalty and/or less than maximal inspection probability (as was in the case without noise). In sum, Proposition 4 is robust to the introduction of noise.

4 Drawing out the Implications

It seems obvious, and implicit in the design of actual policy regimes, that optimal policy towards whistleblowing be sensitive to the motivation attributed to whistleblowers. Our analysis makes the relationship explicit. We find that the optimal

policy involves ‘full enforcement’ – that is, pursuing every case brought to light by whistleblowers and the use of maximal penalties – only when whistleblowing is a social act (our Assumption β) and when the whistleblowing is not distorted by noisy information. We could consider this the benchmark case. In circumstances where whistleblowing is motivated by personal conscience cleansing (Assumption α), or if there is reason to doubt the reliability of the whistleblowers’ information, optimal policy calls for less than complete responsiveness but yet the case for maximal penalty is preserved. However, where one suspects that whistleblowing is punitive in intent (Assumption γ), even the use of maximal penalty may not be optimal as in this case large penalties induce the wrong sort of whistleblower to come forward.

Significantly, optimal policy involves routine pursuit of whistleblower cases only under Assumption β . Under the assumption probably most preponderant in the research literature, α , the optimal probability of pursuit is strictly less than one, even when enforcement is not costly.

There are two ways of thinking about the cases. Firstly, one can see the three assumptions as representing alternative, mutually exclusive theories of what drives whistleblowing. It is then a case of deciding which theory you subscribe to and designing the optimal policy accordingly.

Alternatively, and perhaps more usefully, one can think of the predominant motives for whistleblower activity as likely to vary across settings depending upon local ‘culture’. Those variations might be geographical (countries, for example), but we might also think of mentalities differing between public and private sector, for example.¹⁵

It is sometimes suggested that the kind of people who self-select themselves into public sector employment are endowed with ‘public service motivation’ (Brewer and Selden (1998)). If – and this is a strong if – one believes that whistleblowers drawn from this population are likely to correspond to our Assumptions α or β , our theory

makes a case for policy regimes to be more fully responsive than otherwise.

A small number of studies have investigated the attitudes towards whistleblowing across cultures and found substantial variation. These are generally based on survey responses to hypothetical ethical situations. For example, Keenan and Remington (2002) find that Philippine managers have a ‘flatter’ evaluation of employer wrongdoing (measured along the dimension ‘harm to others’) than their American counterparts – they are more likely to report comparatively minor instances of wrong-doing, but less likely to report the comparatively major cases. Tavakoli et al (2003) present evidence of differences between American and Croatian managers, and Keenan (2002) between Americans and Indians. Patel (2003) investigates generic attitudes towards whistleblowing amongst professionals in Australia (which he takes as a proxy for a cluster of countries including the US, UK and Canada), China and India. He shows the Australians, other things equal, to be “... both *more likely* and *more accepting* of engaging in whistle-blowing.” (Patel (2003: 69), italics in original).

There is a much more substantial body of empirical literature that evidences cross-cultural variation in business ethics more generally (see Jackson and Artola (1997) for a survey). Though not specifically about whistleblowing Becker and Fritsche (1987) found that “(T)he Danish sample of managers would be far more likely to approve the release of pollution into the air if requested by their senior manager than the sample from the US and New Zealand.” Meanwhile Lyonski and Gaidis (1991) found that, across a range of behaviors, “(T)he Danish and German (and, to a lesser extent, the French) sample would be more likely to behave in what might be regarded as a pragmatic way, with Americans more likely to take a principled approach.” In our setting the pragmatic approach might suggest a preference for Assumption β , where the whistleblower has regard to the likelihood that his actions will make a difference, over Assumption α which is more idealist in spirit. That would point (other things equal) towards a more receptive approach to whistleblow-

ers in Denmark and Germany than the US. Even within a country moral judgements on business practices can vary widely. Redfern and Crawford (2004) found a significant association between modernization and judgements of ethicality in various scenarios of business malpractice: “(M)anagers from the more industrialized and economically advanced provinces in China gave harsher moral judgements, that is higher ratings of unethicity, in five out of six vignettes,” (Redfern and Crawford (2004: 48)).

A number of scholars have also explored the link between religion and religiosity on whistleblowing. The results of Barnett et al (1996: 1161) “...indicated that religiosity was positively associated with an ethical ideology of non-relativism. Individuals whose ethical ideologies could be described as idealistic (non-relativist) were more likely to report the wrong-doing of peers.” Again, the non-relativist attitudes might lead us to favor Assumption α and the associated policy conclusions. Pava (1998) identifies a set of distinctively Jewish business ethics, while Park et al (2005) emphasize the Confucianist basis for Korean business ethics.

The cross-cultural variation in attitudes should lead us to suspect some diversity in optimal whistleblower policy. Potentially this could provide the basis for some positive analysis – one could, for example, try to explain variation in whistleblower policy across jurisdictions by variations in ethical attitudes. Such work might be regarded as premature at this stage, however. Whistleblower policy is a rapidly evolving area and it is only in recent years have whistleblowers have moved into such a central role in a number regulatory settings. It would be heroic to think that the current state of law and practice in this area would yet have converged to the optimum, even if one were to accept that the forces of policy development lead to such convergence. The reality that is clear from policy commentary is that most regulatory agencies are still getting to grips with the problem of how to build whistleblowers into their broader enforcement processes and practices.

Existing empirical analyses are somewhat limited for our purposes. None of the

empirical studies cited here ask questions in quite the way that would be needed to differentiate amongst our three theories (though some of the results are suggestive). None of them, for example, told survey respondents the implications for their employer of prosecution. So data would need to be constructed. It is straight-forward to envisage how that could be done by exposing survey respondents or experimental subjects to a set of scenarios designed systematically to vary both the characteristics of the behavior (c and d in our model) and the enforcement environment (π and f).

More fundamentally, the variations in ‘whistleblower policy’ that are typically observable across jurisdictions relate to robustness of whistleblower protections. But what we are modeling is the probability that a given tip will be pursued. This is less likely to be observed (for example because enforcement agencies will not necessarily announce the reason why they decide to inspect plant X) and is embodied in agency practice, budgeting processes and so on.¹⁶ Even if data were available on number of tips received, and the number pursued, we would only be observing equilibrium outcomes rather than propensities.¹⁷

Despite the main ambitions being normative – to provide some guide to the characterization of optimal policy – we hope that future empirical work will get to grips with some of the practical problems identified. It may be that experimental methods would be particularly well-suited here.

5 Conclusions

Manipulating the basic model under three alternative assumptions about why employees become whistleblowers meant we could avoid having to make a definitive choice in a context where the informing literature lacks consensus. Though the model is particular, the analysis points to two general lessons. Both counsel that in designing an enforcement regime informed by the reports of whistleblowers, care needs to be taken to be clear about whistleblower motives.

First, the value of the information that whistleblowers bring to the enforcement agency – and what the agency will wish to do with that information – depends upon the motives ascribed to whistleblowers. Second, in adjusting the enforcement instruments – π and f in our model – attention has to be paid to the change induced in the flow of disclosures, in addition to the direct effect on compliance incentives. Again, the quantitative and qualitative response will depend upon whistleblower motives. Our qualitative findings are robust to the possibility that whistleblowers be less than perfectly informed about the circumstances that prompt firms to violate.¹⁸

One might consider a variety of extensions. In a repeated setting a firm’s past performance, for example, may matter, though in ways that may not be straightforward to predict. An individual may be able to ‘forgive’ or turn a blind eye to wrongdoing on a single occasion, but not when it becomes a pattern. Alternatively, individuals working in organizations with a long record of good behavior may come to regard the firm as implicitly contracted to continue working in that way. In that case a non-compliance decision may be perceived by an employee as a betrayal, and ‘betrayal aversion’ or the related ‘betrayal heuristic’ (Sunstein (2003)) may make disclosure more likely. One could also consider employers differentiated along more dimensions than simply cost of compliance, or consider a budget-constrained or alternatively-motivated regulatory agency.

There are a number of other directions in which model here could be developed. Two that are priorities in future work are, (a) to explore the role played by whistleblower rewards or ‘bounties’ and (b), to investigate how whistleblower-informed inspections could/should be combined with other enforcement instruments, such as random inspections.

References

Alford, C. Fred. 2001. *Whistleblowers*. Ithaca, NY: Cornell University Press.

Andreoni, James. 1990. "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving?," 100 *Economic Journal* 464-77.

Apestequia, Jose, Martin Dufwenberg, and Reinhard Selten. 2007. "Blowing the Whistle," 31 *Economic Theory* 143-66.

Barnett, T., K. Bass, and G. Brown. 1996. "Religiosity, Ethical Ideology, and Intentions to Report a Peer's Wrongdoing," 15 *Journal of Business Ethics* 1161-1174.

Becker, Gary S. 1968. "Crime and Punishment: An Economic Approach," 76 *Journal of Political Economy* 169-217.

Becker, H. and D. Fritsche. 1987. "Business Ethics: A Cross-Cultural Comparison of Managers' Attitudes," 6 *Journal of Business Ethics* 289-295.

Brewer, Gene A. and Sally Coleman Selden. 1998. "Whistle Blowers in the Federal Civil Service: New Evidence of the Public Service Ethic," 8 *Journal of Public Administration Research Theory* 413-440.

Fehr, Ernst and A. Falk. 2002. "Psychological Foundations of Incentives", 46 *European Economic Review* 687-724.

Fehr, Ernst and Klaus M. Schmidt. 1999. "A Theory of Fairness, Competition and Cooperation," 114 *Quarterly Journal of Economics* 817-68.

Glazer, Myron P. and Penina Glazer. 1989. *The Whistleblowers*. New York: Basic Books.

Groeneweg, Sheryl. 2001. "Three Whistleblower Protection Models: A Comparative Analysis of Legislation in Australia, the United States and the United Kingdom," Research Directorate, Public Service Commission of Canada: Ottawa.

Heyes, Anthony. 1996. "Cutting Environmental Penalties to Increase Compliance," 60 *Journal of Public Economics* 251-65.

—. 2000. "Implementing Environmental Regulation: Enforcement and Compliance," 17 *Journal of Regulatory Economics* 107-29.

Hunt, Geoffrey. 1997. "Whistleblowing" in *The Encyclopedia of Applied Ethics*. New York: Academic Press.

Jackson, Terence and Marian C. Artola. 1997. "Ethical Beliefs and Management Behaviour: A Cross-Cultural Comparison," 16 *Journal of Business Ethics* 1163-73.

Jos, Philip, Mark Tompkins, and Steven Hays. 1989. "In Praise of Difficult People: A Portrait of the Committed Whistleblower," 49 *Public Administration Review* 552-61.

Kalb, Paul E. 2000. "Health Care Fraud and Abuse," 282 *Journal of the American Medical Association* 1163-68.

Keenan, John P. 2002. "Comparing Indian and American Managers on Whistleblowing," 14 *Employee Responsibilities and Rights Journal* 79-89.

Keenan, John P. and Steven Remington. 2002. "Cultural Influences on Whistleblowing: A Study of Philippine and US Managers," Proceedings of the Thirty-First Annual Meeting of the Western Decisions Sciences Institute, April 2-5, 2002, Las Vegas, Nevada.

Lawley, Erin. 2007. "With New Regulations, Medicaid Whistleblower Cases Could See Dramatic Increase," *Nashville Business Journal*, March 2, 1.

Lieberman, Stuart. 2007. "Whistleblowers Can Prevent Toxic Nightmares," *International Real Estate Digest*.

Lifton, Robert J. 1986. *The Nazi Doctors: Medical Killing and the Psychology of Genocide*, Cambridge MA: Harvard University Press

Lyonski, Stephen and William Gaidis. 1991. "A Cross-Cultural. Comparison of the Ethics of Business Students," 10 *Journal of Business Ethics* 141-150.

Miceli, Marcia and Janet Near 1992. *Blowing the Whistle: The Organizational and Legal Implications for Companies and Employees*. New York: Lexington Books.

Organisation for Economic Co-operation and Development (OECD). 2004. *OECD Principles of Corporate Governance (Revised)*. Paris: OECD.

Park, Heungsik, Michael T. Rehg, and Donggi Lee. 2005. "The Influence of Confucian Ethics and Collectivism on Whistleblowing Intentions: A Study of South Korean Public Employees," 58 *Journal of Business Ethics* 387-403.

Patel, Christopher. 2003. "Some Cross Cultural Evidence on Whistleblowing," 22 *Journal of International Accounting Research* 69-96.

Pava, Moses. 1998. "The Substance of Jewish Business Ethics," 17 *Journal of Business Ethics* 603-17.

Prendergast, Canice. 2003. "The Limits of Bureaucratic Efficiency," 111 *Journal of Political Economy* 929-58.

Redfern, Kylie and John P. Crawford. 2004. "An Empirical Investigation of the Influence of Modernisation on the Moral Judgements of Managers in the Peoples Republic of China," 11 *Cross Cultural Management* 48-61.

Sunstein, Cass R. 2003. "Moral Heuristics", John M. Olin Law & Economics Working Paper #180, University of Chicago Law School.

Tavakoli, A, John P. Keenan and Cranjak-Karanovic B. 2003. "Culture and Whistleblowing An Empirical Study of Croatian and United States Managers," 43 *Journal of Business Ethics* 49-64.

United States Attorney's Office, Western District of Washington. 2007. "Shipping Line Pays \$750,000 in Fines and Assessments for Illegal Ocean Dumping," posted at www.usdoj.gov/usao/waw/press/2007/jan/irika.html.

US Department of Labor. 2005. *FY 2005 Performance Budget Overview*, available online at www.dol.gov/_sec/Budget2005/overview-pb.htm.

Vyverberg, Robert W. 2001. "Unplug Your Ears: Whistleblowing Is Alive and Well," 11 *Employment Law Letter*, 1-2 (May).

Watcher, The. 2003. "Bush Administration Ignores Whistleblowers," July 28, 2.

Notes

¹Vyverberg (2001) and Kalb (2000) provide examples. Notably, Cooper and Wadkins (along with Coleen Rowley, the FBI agent who exposed her employer's slow reaction to intelligence prior to 9/11) appeared on the front cover of *Time* as the magazine's *People of the Year* in 2003. Hamel was the subject of the 2004 TV documentary *The Whistleblower*.

²Examples of legislation that prohibits discharging or discriminating against an employee who reports an actual or anticipated violation of the Act include the Clean Air Act Amendments (1989), the Financial Institutions Reform, Recovery and Enforcement Act (1989), the Asbestos Emergency Response Act (1996), and the Pipeline Safety Improvement Act (2002).

³Whistleblowing is further encouraged by the establishment of programs that allow anonymous reporting of wrongdoing. See, for example, the UK Financial Services Authority's recently established "Whistleblower Line" and similar programs of the Government Accountability Project in the US.

⁴We do not consider here the optimal level of protection given to whistleblowers, taking as given that an effective regime of whistleblower protection is in place. The 'how responsive' question can also be thought of as relating to the funding of these programs. If funding of whistleblower programs is limited, or must compete with other activities for resources, the agency will be obliged to ration its follow-up activity. In 2005, for example, the US Department of Labor bid for an extra \$2m "... to improve whistleblower investigation" (US Department of Labor 2005). Whistleblower groups often criticize agencies for slow or inadequate response to issues raised by whistleblowers. In the context of public employees, for example, "...though employees have been blowing the whistle more and more in recent years, the Bush administration does not appear to be listening. Law requires the Office

of Special Counsel (OSC) to resolve cases within 15 working days; however most of the reports in the backlog have sat for more than 6 months, and at the beginning of June 2003 OSC had a backlog of 628 whistleblower allegations awaiting review” (The Watcher 2003).

⁵Apesteguia *et al.* (2007) examine the role of policies that encourage a cartel member to inform a regulator about the wrongdoing of fellow cartel members in exchange for a reward. This is quite different to an agent (employee) blowing the whistle on a principal (employer) which is what we consider here. Outside economics there is an extensive literature on whistleblowing, mainly in the areas of sociology, psychology, business and public administration: Miceli and Near (1992) provide a good survey. The assumptions of our model are informed by this body of research.

⁶Wikipedia defines a whistleblower as “an employee or member of an organization ... who reports misconduct to people or entities that have the power to take corrective action.” It goes on to refer to the prevention of “direct threats to public interest – fraud, safety violations, and corruption are just a few examples.”

⁷In recent years, and in some contexts (especially financial fraud), US courts have approved the paying of ‘bounties’ to whistleblowers whose revelations lead to successful prosecution. We do not incorporate bounties into our model, though we note in the conclusions that this would be a natural and analytically straightforward extension. Insofar as the motive for passing information to enforcement agencies became mercenary it would cease to be whistleblowing.

⁸There are alternative assumptions that we could have made. The regulator could, for instance, have been entrusted with a ‘stewardship’ role and so act to minimize expected external damage. But welfare maximization is a common approach, and allows us to focus on optimal policy.

⁹Alford’s theory of ‘narcissism moralized’ is closely related to sense of shame.

“Though not all whistleblowers use the language of shame, many talk about feeling dirty or corrupted by the acts of others with whom they are associated, and an inability to live with that corruption” (p. 74). The psychological analysis of the need to conscience cleanse is related to the cognitive process of ‘doubling’ (Lifton (1986)), and thereafter to mental disfunction.

¹⁰It also bears a family resemblance to the idea that people’s taste with regard to retribution are driven by an ‘outrage heuristic’: “People’s intuitions about punishment seem quite disconnected from the consequences of punishment. People want to punish the corporation as if it were a person. Punishment judgements are rooted in a heuristic, to the effect that penalties should be a proportional to the outrageousness of the act” (Sunstein (2003:5)). Significantly outrageousness of an act may not be closely correlated with the forward-looking costs and benefits of intervention. Sunstein offers experimental evidence that subjects levy the same punitive damages in hypothetical contexts in which though the past violation was common, the next period implications of penalty levels differ.

¹¹The construct is best understood by contemplating the extremes. If my employer chose to not comply despite having a c equal to zero then that might be a particularly outrageous or indefensible decision – showing an unwillingness to act to protect the interests of others, even though such action would be personally costless – and quite likely to prompt disclosure. As c gets larger it is more defensible that the firm would opt to violate, though individuals may differ in their ‘forgiveness’. At the upper extreme, as c approaches infinity, then non-compliance would be much more ethically excusable - indeed unavoidable.

¹²It might also embody any private costs associated with whistleblowing, that we have excluded here.

¹³A formal proof is by contradiction. Consider any regime $\{\pi', f'\}$ such that

$\widehat{c}(\pi', f'|\alpha) \geq d$. Differentiating (5) gives:

$$\frac{\partial SL(\pi', f'|\alpha)}{\partial \pi} = \frac{\partial \widehat{c}}{\partial \pi} \cdot (1 - \rho_\alpha(\widehat{c})\pi)(\widehat{c} - d) + \int_{\widehat{c}}^{\infty} \rho_\alpha(c_i)(c_i - d)dG(c).$$

The first term on the right-hand side is non-negative and the second is positive implying $SL(\pi', f'|\alpha) > SL(\pi' - \epsilon, f'|\alpha)$, such that $\{\pi', f'\}$ cannot be optimal, contradicting our assumption.

¹⁴Formally, in the relevant first-order condition

$$\frac{\partial SL}{\partial \pi} = \frac{\partial \widehat{c}}{\partial \pi} [(1 - \rho_\beta(\widehat{c})\pi)(\widehat{c} - d)] + \int \int_{\widehat{c}}^{\infty} \rho_\beta(c|\varepsilon)(c - d)dG(c)dE(\varepsilon)$$

the first term is negative but the second is ambiguous in sign: if so, an interior maximum for π cannot be ruled out.

¹⁵In any actual population, however, there would be a variety of motives for why individuals report planned wrong-doing in some contexts but not others. Envisaging a population that is a mixture of α -, β - and γ -types in some known proportions leads to policy implications that can be thought of as ‘weighted averages’ of the pure cases. Allowing for mixed motives complicates the analysis without adding much to our understanding.

¹⁶A variety of other data problems arise in practical settings. For example, an observer would not be able to distinguish those whistleblower cases dismissed because on a preliminary screen they are regarded as off-the-mark, or not furnishing sufficient detail or evidence, from those that fall within the $(1 - \pi)$ fraction. The former type do not feature in our model, but surely do in real world settings.

¹⁷For example, if we observed that a particular enforcement agency received 100 tips and pursued 50 of them, the realized (equilibrium) probability of pursuit equals 0.5. This does not, though, imply that if the number of tips were to increase to 200 the agency would necessarily pursue 100 of them. That would rely on, amongst other things, the budgeting process – it may be that the office only has enough

resource to pursue 50 cases per period. So π – the crucial policy variable in our analysis – would take some ingenuity to uncover.

¹⁸There are various ways in which ‘noise’ can be introduced into our framework. One could investigate the potential role of strategic behavior or lying on the part of whistleblowers, or the implications of potential for error in the evidence chain, or of slippage or imperfection in the enforcement process itself. Whilst each of these would render the model inconsistent with the strict definition of whistleblowing that we have adopted, they would nevertheless be useful directions for further research.

Figure 1: Relationship between compliance cost and probability of disclosure under conscience cleansing

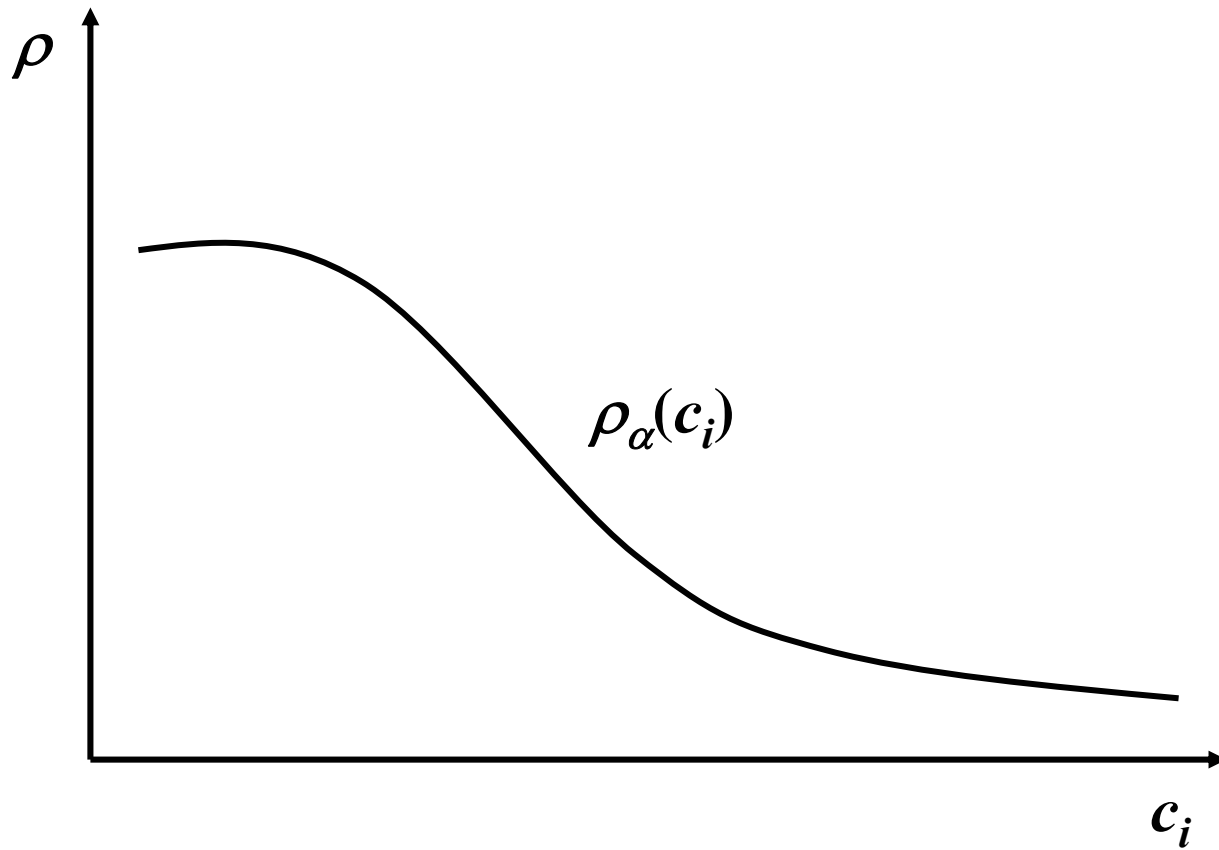


Figure 2: Realized probability of compliance (solid line) under optimal policy *versus* first-best (dashed)

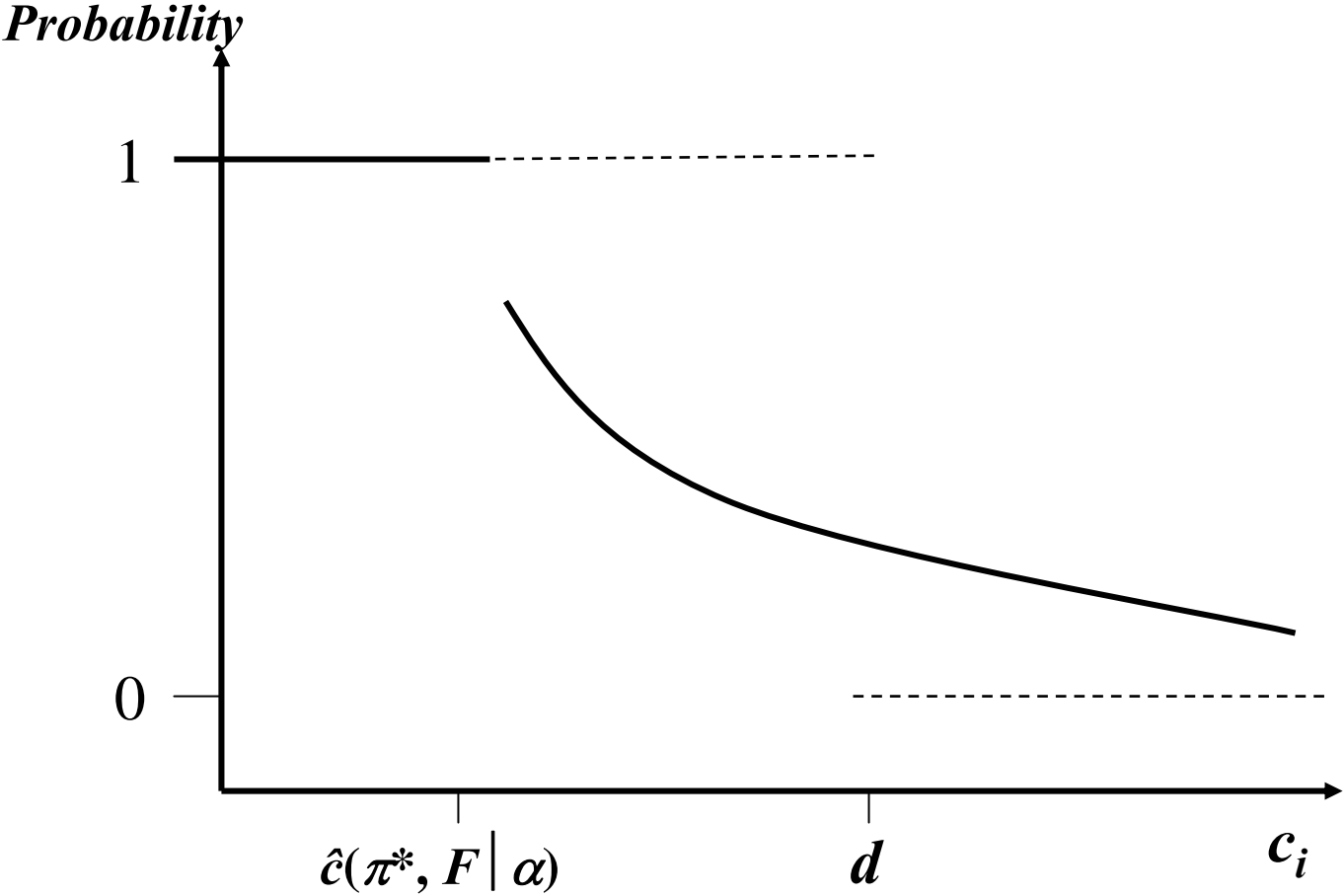


Figure 3: Realized probability of compliance in example

