doi:10.1017/S0003055408080192

Whistleblowing

MICHAEL M. TING *Columbia University*

By skipping managers and appealing directly to politicians, whistleblowers can play a critical role in revealing organizational information. However, the protection of whistleblowers can affect managers' abilities to provide employees with incentives to exert effort. This paper explores this tradeoff with a model of agency decision-making under incomplete information. In the game, an employee's effort determines a project's quality, and a manager chooses whether to approve the project and discipline the employee. The employee and politician wish for only "good" projects to be approved. By whistleblowing, an employee reveals the quality to a politician outside of the organization, who may override the manager's decision. A key finding is that from the politician's perspective, the benefits of whistleblower protections depend on the preferences of the manager. If the manager is inclined toward approving projects, then the costs of lower employee effort may outweigh the informational benefits of whistleblowing. The optimal policy may then be to ban whistleblowing. By contrast, when the manager is inclined toward rejecting projects, whistleblower protections prevent him or her from suppressing effort and are unambiguously beneficial.

Whistleblowers have historically played key roles in passing crucial information from lower levels of organizations to higher level officials. A casual survey of American organizations in recent years amply demonstrates that this trend has not abated. In 2002, Federal Bureau of Investigation (FBI) staff attorney Coleen Rowley went public over the bureau's investigation of the alleged 9/11 co-conspirator Zacarias Moussaoui. Her account of how FBI headquarters stifled attempts to investigate his activities built support for the reorganization of its antiterrorism efforts. In 2004, Food and Drug Administration (FDA) researcher David Graham testified before a Senate committee that the agency had ignored warnings about the heart disease risks posed by Vioxx prior to its approval. These revelations caused serious damage to the FDA's credibility and generated demand for both stricter drug approval procedures and improved postapproval monitoring. Such episodes have not been confined to the public sector. In 2002, Sherron Watkins of Enron and Cynthia Cooper of WorldCom both gained acclaim for their roles in uncovering managerial irregularities in their respective corporations.[1]

Coincident with its practice, whistleblowing has long enjoyed political and legal protection. In the United States, rudimentary protections were first enacted by the Continental Congress. A centerpiece of the modern legal framework dates to 1863, when Congress passed the False Claims Act (FCA) in order to combat Civil War profiteers. The law allowed citizens—termed *qui tam* relators—to bring a suit against an alleged offender on behalf of the government, and to share in a percentage of the damages awarded. More recent legislation has focused on the relationship between employees and management. In 1978, the Civil Service Reform Act criminalized retaliation against whistleblowers and created procedures for reversing terminations of their employment. The Whistleblower Protection Act (WPA) of 1989 promised confidentiality of whistleblower disclosures and further limited the ability of managers to retaliate against employees.[2] These laws have been amended (and usually strengthened) numerous times. Other protections are in place outside the federal government. Most U.S. states have enacted similar laws, and courts have frequently protected whistleblowers even in the absence of explicit protections. Private-sector whistleblowers are also protected to varying degrees by federal and state laws, such as the 2002 Sarbanes–Oxley Act.[3]

There exists something of a consensus today that the legal protection of whistleblowing, like many bureaucratic reforms dating to the Progressive Era, is an essential part of effective government.[4] The institutional logic thereof is typically based on two fairly innocuous observations. First, the very idea of bureaucratic organization suggests that an agency's principals cannot specify ex ante all the actions that it should take; that is, some actions are uncontractable. Second, principals

Michael M. Ting is Associate Professor of Political Science and Public Affairs, Political Science Department, Columbia University, 420 W 118th St., New York, NY 10027 (mmt2033@columbia.edu).

This research is generously supported by National Science Foundation Grant SES-0519082. This paper has benefited tremendously from the input of seminar audiences at MIT, Princeton, the University of North Carolina at Chapel Hill, Washington University, the University of Chicago, and Berkeley. I thank David Austen-Smith, Tom Hammond, Greg Huber, Stuart Jordan, George Krause, Patrick Warren, and Alan Wiseman for helpful comments; and Arne Grafweg, Caroline McGregor, and Andrea Venneri for research assistance.

[1] Watkins and Cooper shared, along with Rowley, *Time* magazine's 2002 People of the Year award as "The Whistleblowers."

[2] Disclosures are typically handled by some combination of the employing agency's Inspector General, the Office of Special Counsel, and the Merit Systems Protections Board. Other federal whistleblowing laws are implemented by relevant regulatory agencies, for instance, the Occupational Safety and Health Administration.

[3] Since 1999, whistleblower protection laws have also been enacted in Australia, the UK, New Zealand, South Africa, and Canada. See Lewis (2001) for a comparative assessment.

[4] As an example of the prevailing normative orientation toward whistleblowing, Shafritz and Russell's (2000) *Introducing Public Administration* defines "whistleblower" as "[a]n individual who believes the public interest overrides the interests of his or her organization and publicly blows the whistle on—meaning exposes—corrupt, illegal, fraudulent, or harmful activity." See also De Maria (1999) and Alford (2001).

https://doi.org/10.1017/S0003055408080192 Published online by Cambridge University Press

(such as Congress) can therefore benefit from the information possessed by organization members (such as research staff) who do not normally interact with them.

This paper argues that the prevailing logic is incomplete. Whistleblower protections such as reducing the scope of managerial punishments or allowing employees to claim some of the manager's surplus are certainly desirable to a principal in some circumstances. But in others they may undermine an agency's output by diluting incentives for employee effort. Thus it is not always optimal for a principal to reduce the cost of blowing the whistle or to reduce managerial authority to discipline employees. In some cases, he or she may even wish to ban whistleblowing altogether.

To see the intuition for these results, consider a manager who is charged with approving or rejecting a project, such as a license, contract, investigation, pharmaceutical application, or rocket launch. The project may be of high or low quality, and a politician overseeing an agency and an employee working in it both want the high-quality project approved and the low-quality project rejected. Without information about quality, the politician may be exploited by an "aggressive" manager who wants to approve the project regardless of quality. In this environment the employee can reveal low quality (i.e., blow the whistle), and thereby inform the politician that the manager should be overruled.

Now suppose that instead of being fixed, quality is determined by employee effort, with higher effort making high quality more likely. The manager therefore wishes to provide incentives for the employee to work hard, because all players would agree to the approval of a high-quality project. However, the possibility of whistleblowing changes this calculation. Instead of using her punishment capacity exclusively to induce higher effort, the manager may now divert it toward deterring whistleblowing. Further, an employee who can whistleblow may not face the full consequences of a bad project, because the principal will replace an exposed manager. Both effects lower the power of the employee's incentives and hence make low quality more likely. Thus, in the presence of an aggressive manager, the principal may be better served by discouraging whistleblowing.

This logic is reversed when the manager is "conservative" or wants to reject the project regardless of quality. This manager attempts to inhibit employee effort, because the principal would agree only with a rejection of a low-quality project. Whistleblowing now *raises* employee effort, because it raises the employee's return to effort by ensuring that a high quality project will be revealed and approved once the manager is replaced. Equivalently, it reduces the manager's ability to induce the employee to bring about a project that all players would want cancelled. A principal therefore unambiguously desires stronger whistleblowing protections.

These examples highlight a tension between two kinds of organizational incentives. *Ex post*, there are incentives to reveal information, and *ex ante* there are incentives to exert effort, which affects the information to be revealed. Arguments for strengthened whistleblower protections often focus on the former: once quality is determined, the employee and politician must benefit from whistleblowing. A classic intuition of organization theory focuses on the latter. For decades, it has been argued that organizations must maintain a "chain of command," whereby subordinates report only to immediate superiors (e.g., Fayol 1949; Bolton and Dewatripont 1994). Among other rationales, the prohibition of "skip-level" reporting improves performance by removing perverse managerial incentives. A manager worried about being publicly exposed by a subordinate might divert effort toward suppressing employees, or select bad employees (e.g., Friebel and Raith 2004).[5]

The model developed here exploits these incentives to characterize the strategies of politicians, managers, and employees in the presence of possible whistleblowing. It generates predictions about when whistleblowing will occur, as well as its effects on organizational performance. These in turn will determine the level at which a politician might implement protections. Thus in addition to developing the logic of an important mechanism for communication between the bureaucracy and its principals, the model may help to explain some of the historical variation observed in whistleblowing laws.[6]

The game has three players; a manager and an employee who form an organization, as well as a principal who monitors its behavior. The players contribute to the output of a project that generates a publicly observable outcome in each of two periods. All players are interested in these outcomes, but differ in knowledge and ability. Organization members are also motivated by the possibility of sanctions from higher levels of the hierarchy. This environment perhaps best describes a public agency, where the manager is a political appointee; and the employee, a civil servant. Thus the manager is relatively easily replaceable, whereas the employee's terms of employment cannot be altered easily, and civil service rules heavily constrain the incentives that managers can provide (e.g., Knott and Miller 1987). To a lesser degree, it may also apply to voters and elected officials or to shareholders and

[5] Several significant court cases have invoked the rough intuition of the tension between *ex ante* and *ex post* incentives. Most prominently, in the controversial May 2006 *Garcetti v. Ceballos* decision, the Supreme Court held that statements by government employees do not enjoy First Amendment protection from managerial discipline. In the court's opinion, Justice Anthony Kennedy wrote that "Government employers, like private employers, need a significant degree of control over their employees' words and actions; without it, there would be little chance for the efficient provision of public services." In dissent, Justice David Souter argued that "private and public interests in addressing official wrongdoing and threats to health and safety can outweigh the government's stake in the efficient implementation of policy." See U.S. Supreme Court docket 04-473.

[6] These variations include shifts in the awards given to *qui tam* relators, as well as redefinitions of the agencies and activities covered by the WPA. Most importantly, 1994 amendments to the law significantly increased the range of protected disclosures. However, the scope of the WPA has often been reduced by court decisions. In response, the House of Representatives in March 2007 passed H.R. 985, the Whistleblower Protection Enhancement Act, which would expand protected disclosures and extend WPA protections to national security agencies. As of this writing, the House and Senate had not agreed on a compromise. See Congressional Research Service report RL33918 (2007) for an overview of the WPA.

https://doi.org/10.1017/S0003055408080192 Published online by Cambridge University Press

management in public corporations. Principals in these environments have some ability to replace managers but have less ability to choose the institutional context of whistleblowing.

As the examples suggest, the amount of protection optimally afforded to whistleblowers depends centrally on managerial preferences and, in particular, on the relative inclination toward approving the project instead of rejecting it and allowing a status quo project to prevail. Managerial preferences fall into two categories of interest. In both, the manager reflects the principal's preferences imperfectly (perhaps because of separation of powers), thereby creating the possibility that the principal would wish to overrule her decision. An aggressive manager wishes to approve projects that the principal would not. Such a manager might, for example, be more inclined to override safety concerns in rocket launches or pharmaceutical applications. By contrast, a conservative manager wishes to reject projects that the principal would not. This manager might be hesitant to follow up on investigative leads. A convenient shorthand is therefore to think of an aggressive manager as erring on the side of Type I errors, whereas a conservative manager errs on the side of Type II errors.

The game begins with the employee's choice of a costly and nonverifiable effort level. This effort probabilistically determines the project's quality, which is initially observable only to the employee and manager. In the first period, the manager chooses whether to approve the project. If the project is approved, then an outcome correlated with its quality is generated. Between periods, the manager and employee are each able to reveal project quality to the politician. In this context, whistleblowing is simply an employee report. To reflect the sometimes considerable burden of coming forward as well as the effects of whistleblower protection laws, this report may be costly for the employee. This report is "hard," or verifiable, information such as expert testimony, which can only be provided voluntarily by organization members. Following the report(s), the manager can punish the employee. Although civil service protections restrict the extent of managerial discretion over employee rewards, these incentives can plausibly include task assignments, performance reviews, or other benefits of office. This punishment is costless to the manager, who may thus effectively commit to a punishment schedule as if it were a contract. Finally, the politician chooses whether to revoke the manager's decision rights in the second period, or allow him or her to continue exercising approval authority. Intuitively, revocation places the manager's department in receivership. Thus whistleblowing renders the managerial decision contractable.

The model's key innovations are that it considers simultaneously an agency's internal structure and external environment, as well as the interaction between effort and information. Naturally, the model also has a number of limitations. First, it does not feature a spatial policy dimension and instead focuses on the implementation (or nonimplementation) of a program pursuant to some established law. Second, it ignores many transaction costs (although it includes a personal cost of blowing the whistle). One argument against elaborate whistleblower protections is that they impose nontrivial costs on various actors (e.g., Kovacic 1998), and so the model is presumptively favorable to such protections. Third, because there is no repeated play, it is silent on the effect of whistleblowing on norms arising from reputation or culture. Finally, it is concerned generally with organizational performance, and not with the protection of whistleblower rights per se.[7]

To date, there has been relatively little scrutiny of the effects of whistleblowing on organizational performance.[8] There exists a small and recent industrial organization literature that views whistleblowing as defection from collusive behavior amongst peers in a repeated game setting. For example, Aubert, Rey, and Kovacic (2006) predict that incentives such as prosecutorial leniency for offending firms will undermine cartels. On the other hand, Apesteguia, Dufwenberg, and Selten (2007) provide mixed experimental evidence for the effectiveness of such incentives.

The model's three-tier structure complements an influential body of three-tier principal-agent models. These models compare the performance of three-tier hierarchies against alternative organizational forms (e.g., McAfee and McMillan 1995; Melumad, Mookherjee, and Reichelstein 1995), or address the possibility of collusion between organization members against the principal (e.g., Tirole 1986; Laffont 1990). By contrast, whistleblowing describes the situation opposite to that of collusion: organizational players in effect compete to influence the principal. A notable and somewhat related exception is Prendergast (2003), who develops a model of an organization's allocational efficiency in the presence of customer complaints.

There are presently few political applications of three-tier models. However, this institutional structure is approximated by work on administrative procedures and agency design (McCubbins, Noll, and Weingast 1987; Moe 1989). Although not formalized, these theories examine the rationales for and implications of structures that enfranchise interest groups to participate in agency rulemaking. Under laws such as the Administrative Procedures Act, interest groups can play a whistleblowing role and ensure bureaucratic compliance with legislative wishes. Relatedly, numerous models consider the role of outside actors in helping principals to exercise "fire alarm" oversight over agencies (e.g., McCubbins and Schwartz 1984; Hopenhayn and

[7] Concerns about employee rights do animate much the discussion about whistleblowing policy. For example, a 2001 Canadian report criticized the U.S. system for failure to "focus on the spirit of whistleblower protections" and excessive concern with organizational "efficiency and effectiveness" (Public Service Commission of Canada, 2001). From an organizational performance perspective, these concerns place more weight on *ex post* incentives and less on *ex ante* incentives.

[8] Substantial literatures in law and organizational behavior focus on the legal and ethical dimensions of whistleblowing, as well as the motivations and characteristics of whistleblowers (e.g., Bowman 1983; Near and Miceli 1996).

https://doi.org/10.1017/S0003055408080192 Published online by Cambridge University Press

Lohmann 1996).[9] This body of work is relatively silent, however, on the role groups may play in influencing, as opposed to revealing, policy "type."

Due to its combination of moral hazard and signaling, the model implicitly draws on two significant families of models of bureaucracies. The first considers the provision of incentives within organizations (Gibbons 1998; Dixit 2002; Gailmard and Patty 2007). Of particular relevance are models of multiple tasks (Holmstrom and Milgrom 1981; Ting 2002) and common agency (Dixit 1998; Wilson 2000; Gailmard 2007). In considering an employee that performs two "tasks" (effort and whistleblowing) alongside a manager who effectively faces two principals, the present work integrates both perspectives. Its findings on managerial strategy therefore engages an extensive body of work on personnel policy in the U.S. executive branch (Heclo 1977; Lewis 2003; Krause, Lewis, and Douglas, 2006). The second family addresses the extraction of information from agencies. These models consider a principal's incentives to scrutinize agency reports (e.g., Banks 1989), or her allocation of decision rights (e.g., Epstein and O'Halloran 1994). Typically, however, they do not consider incentive issues within an agency.

The paper proceeds as follows. The next section formally lays out the whistleblowing model. The section under Whistleblowing Strategy derives and discusses the equilibrium of the model. The section Whistleblower Policy then considers implications for whistleblowing policy, including limiting managerial retaliation against whistleblowing and allowing employees to claim part of the manager's surplus. The final section summarizes and concludes.

THE MODEL

The game considers a simple institutional environment with three players: a (P)olitician or principal and an agency or organization composed of an (E)mployee and a (M)anager. There are two periods, indexed where relevant by a subscript t. Players "discount" the second period's payoffs by a factor $\delta > 0$, where I allow $\delta > 1$ to allow the second period to be more important than the first. Thus, the first period may have only a pilot project, while the second has a fully implemented program.

In each period t, the players generate an outcome $x_t \in X \cup \{q\}$, where $X \subset \Re$ is convex and compact. The outcome q can be considered the result a default policy that generates a payoff of zero for all players. If $x_t \neq q$, then player i receives linear payoffs:

$$u^i(x_t) = b^i x_t - k^i, \tag{1}$$

where $b^i > 0$ and $k^i > 0$. Additionally, let $x^i = k^i/b^i$ be the outcome that generates a payoff of 0 (i.e., equal to that of q) for player i. Thus x^i is a "standard" below which player i would prefer q. When the default policy generating outcome q is not chosen, x_t is determined in part by the project's *quality* (or "type") $\theta \in \{\underline{\theta}, \overline{\theta}\}$. Throughout, a project with $\theta = \overline{\theta}$ will be referred to as "high quality," and a project with $\theta = \underline{\theta}$ as "low quality."

Each x_t is drawn i.i.d. according to a continuously differentiable probability density $f(x_t \mid \theta)$ on X. The density functions satisfy the familiar Monotone Likelihood Ratio Property (MLRP):

$$\frac{d}{dx_t}\left[\frac{f(x_t \mid \overline{\theta})}{f(x_t \mid \underline{\theta})}\right] > 0 \tag{2}$$

This ensures that higher observations of x_t are more likely to be associated with high quality. The expected value of x_t is $\overline{x}$ under high quality and $\underline{x}$ under low quality, and so $\overline{x} > \underline{x}$.

The preferences over policy outcomes for each player are usefully defined relative to $\overline{x}$ and $\underline{x}$. To focus attention on the most interesting case, where P cares most about the project's quality, let P prefer the expected outcome of the high quality project to q, and prefer q to the expected outcome of the low quality project:

$$x^P \in (\underline{x}, \overline{x}). \tag{3}$$

Given (3), the exposition will further focus on the two non-trivial cases of managerial and employee preferences. In the first, M is *aggressive* in the sense that $x^M < \underline{x} < x^E < \overline{x} + (\overline{x} - \underline{x})/\delta$. Here M finds both levels of project quality preferable to the default outcome q, although E and P would prefer cancellation of the low-quality project. Note that E may also wish for cancellation of the high-quality project, although her extremism is bounded in order to avoid a few uninteresting corner solutions where she exerts zero effort. In the second, M is *conservative*, with $x^M > \overline{x} > x^E$. Here M wishes to cancel all projects, whereas E and P would prefer approval the high-quality project.

Toghether, these two cases cover a range of common environments of interest. In a separation-of-powers system, a politically appointed manager could easily have preferences that diverge from those of an agency's civil servants and its legislative principals. Thus, an antienvironmentalist executive could appoint a conservative manager to head an environmental protection agency despite a proenvironmentalist legislature.[10] Agency managers may also face structural or interest group pressures that induce aggressiveness or conservatism. For example, changes in the National Aeronautics and Space Administration's launch procedures in the 1980s resulted in a shift toward a more aggressive posture toward launches (Heimann 1993). In the 1990s, pressure from patient advocacy groups helped to generate policies that made the FDA's drug approval process more aggressive (e.g., Carpenter 2004). In both agencies, the changes probably expanded the set of "approvable" projects.

[9] In much of this work, the fire alarm, or transmission of information, is non-strategic. In a somewhat different context, Lorentzen (2007) considers the strategic transmission of information to principals.

[10] The Reagan administration's 1981 appointments of James Watt as Secretary of the Interior and Anne Gorsuch as Environmental Protection Agency Administrator were widely seen as examples of this practice.

https://doi.org/10.1017/S0003055408080192 Published online by Cambridge University Press

It is straightforward to see why these are the only cases in which whistleblowers are of significant interest. If the manager's preferences over approval and cancellation were identical to the principal's (i.e., $\underline{x} < x^M < \bar{x}$), then she would be a perfect agent for the principal and employee whistleblowing would be inconsequential. Similarly, if the manager and employee had identical preferences, then the employee would never have an incentive to whistleblow.

Players inside the agency also receive payoffs from nonpolicy sources. M values office-holding and receives a fixed benefit of $m > 0$ for each period in which she holds managerial control. Additionally, E faces costs from three sources. She can be punished by M, which results in a loss of $p \in [0, \overline{p}]$. This can correspond to reassignment, delayed promotion, or perhaps dismissal. Note that it is crucial that M be able to provide *some* incentives to E. If E possessed managerially relevant information but did not face such incentives, whistleblowing protections would be trivially desirable to P. Next, the act of whistleblowing imposes a personal cost $c_w \geq 0$. Finally, E exerts a one-time effort which affects θ_t. The effort level $e \in [0, 1]$ imposes a cost ce^2, where $c > \max\{0, [(1+\delta)b^E(\bar{x} - \underline{x}) + \overline{p}]/2\}$ to avoid some cumbersome corner solutions.

The game begins with E's choice of e, which is unobservable to M and P. Nature then determines the project's quality, where $\Pr\{\theta = \bar{\theta}\} = e$. The quality is initially observable to E and M but not P. At $t = 1$, the agency then executes the project according to the following sequence:

- M chooses approval decision $a_t \in \{0, 1\}$, where 1 corresponds to approval, and 0 to rejection and $x_t = q$.
- If $a_t = 1$, Nature randomly determines outcome x_t.

The key actions of the model take place between the two project execution stages, after x_1 is revealed. The sequence of this phase is as follows. All actions are observable unless otherwise noted.

- M issues a costless report $r \in \{\emptyset, \theta\}$.
- E makes a *whistleblowing* decision $w \in \{\emptyset, \theta\}$, at cost c_w if $w = \theta$.
- M chooses a punishment level $p \in [0, \overline{p}]$, unobserved by P.
- P chooses period 2 decision rights $s \in \{M, P\}$.

The managerial report and the whistleblowing decision have identical effects. Both either reveal θ fully ($r = \theta$ or $w = \theta$) or convey no additional information ($r = \emptyset$ or $w = \emptyset$) to P. This technology is based on a kind of uncontractability. P does not understand *ex ante* the relationship between θ and a_t. Organization members may "explain" this relationship, but cannot be compelled to do so.

M's punishment is the mechanism through which E receives performance incentives. P's choice of s allows her to punish M by "renegotiating" second period decision-making rights. Note that P may condition s on θ only if $r = \theta$ or $w = \theta$. If $s = M$, then M retains control and the second period of project execution is identical to that of the first. If $s = P$, then P assumes M's role in choosing a_2. P breaks ties in favor of allowing M to retain control. The assumption implicit in this setup is that intervening in management decisions is "difficult" for P. That is, whether due to (unmodeled) labor market constraints, uncontractability, or transaction or opportunity costs, P's primary short-run option in the face of poor performance is the circumvention of management.[11] Thus Congress might react to an adverse agency practice by simply passing a law to ban it.

The solution concept for the game is Perfect Bayesian Equilibrium (PBE) in pure, weakly undominated strategies. Denote by H_1 the set of possible observables (a_1 and x_1) following period 1, and H_2 the set of possible observables prior to period 2. For E, the equilibrium specifies effort $e \in [0, 1]$ and whistleblowing $w : [0, 1] \times \{\underline{\theta}, \bar{\theta}\} \times H_1 \times \{0, \theta\} \to \{\emptyset, \theta\}$ strategies. M's strategy has mappings $a_1 : \{\underline{\theta}, \bar{\theta}\} \to \{0, 1\}$ and $a_2 : \{\underline{\theta}, \bar{\theta}\} \times [0, \overline{p}] \times H_2 \to \{0, 1\}$ specifying period 1 and period 2 approval decisions. It also has measurable mappings $r : \{\underline{\theta}, \bar{\theta}\} \times H_1 \to \{\emptyset, \theta\}$ and $p : \{\underline{\theta}, \bar{\theta}\} \times H_1 \times \{\emptyset, \theta\}^2 \to [0, \overline{p}]$ specifying her reporting and punishment decisions, respectively.

P's strategy consists of the mappings $s : H_1 \times \{\emptyset, \theta\}^2 \to \{M, P\}$ and (abusing notation somewhat) $a_2 : H_2 \to \{0, 1\}$ that assign managerial rights and an approval decision for period 2, respectively. To reduce the number of equilibrium cases, M and P are assumed to break ties in favor of approving projects. Finally, P has posterior beliefs $\mu : H_1 \times \{\emptyset, \theta\}^2 \to [0, 1]$ that $\theta = \bar{\theta}$, given her observation of x_1, r, and w. For discussion purposes, it is useful to define the "intermediate" beliefs $\mu_r : H_1 \to [0, 1]$ and $\mu_w : H_1 \times \{\emptyset, \theta\} \to [0, 1]$ that P holds immediately prior to M's report and E's whistleblowing choice, respectively. If an out of equilibrium information set is reached without θ being revealed, then $\mu = 0$ $(= 1)$ if $x^M < (>)$ $\underline{x}$. These beliefs are "pessimistic" about M's preferred action and incline P toward revoking her authority, but they do not play a significant role in the results.

The model has multiple equilibria, which fortunately do not generally alter the conclusions. To simplify the analysis, two equilibrium selection rules are adopted. The first addresses information revelation strategies. It chooses the "minimum reporting" equilibrium, in which (i) the minimum number of players report θ, and (ii) when either player could reveal θ, M reports when it is in her interest to do so. Part (i) is consistent with the presence of costs in issuing nontrivial reports. It also selects the revelation strategy that maximizes the informed players' equilibrium expected payoffs, as it forces P to give E and M the benefit of the doubt when $r, w = \emptyset$ along the equilibrium path. Part (ii) has the virtue of robustness to "errors" by E, in the sense that M neither relies on E to report information that she would have wanted to reveal unilaterally, nor unilaterally reveals information that would be damaging to her. By the symmetry of the revelation technology, it should be clear that part (ii) cannot affect information

[11] The main results of the model hold (but are more cumbersome) in the presence of positive costs of replacing M.

https://doi.org/10.1017/S0003055408080192 Published online by Cambridge University Press

revelation or equilibrium payoffs (net of c_w). The second addresses effort levels, by choosing the equilibrium in which E chooses the highest effort level. This yields the optimal equilibrium for both E and P. The effects of these rules are discussed in the following section.

WHISTLEBLOWING STRATEGY

This section discusses strategies and presents the main results on effort levels and punishments in Propositions 1–2. It is helpful to begin by adopting the following terminology about project quality and beliefs. A quality level θ is *aligned* if the manager and principal agree on the proper approval decision. Thus high quality ($\theta = \overline{\theta}$) is aligned under an aggressive manager, and low quality ($\theta = \underline{\theta}$) is aligned under a conservative manager. The manager likes θ to be aligned because this removes the principal's incentive to "fire" her before period 2.

Next, I introduce a notion of "correctness" of the principal's beliefs. Given a belief μ that the project is of high quality, the principal is indifferent between approval and rejection if $b^P(\mu\overline{x} + (1-\mu)\underline{x}) - k^P = 0$. Solving yields the following threshold value of μ:

$$\tilde{\mu} = \frac{k^P/b^P - \underline{x}}{\overline{x} - \underline{x}}. \tag{4}$$

Note that by (3), $\tilde{\mu} \in (0,1)$. Now $\mu < \tilde{\mu}$ implies pessimism by the principal of high quality, who would then want to reject the project, whereas $\mu > \tilde{\mu}$ implies optimism and approval. The principal's beliefs are then *consistent* if $\mu < \tilde{\mu}$ and quality is low, or if $\mu \geq \tilde{\mu}$ and quality is high. Either a nonempty managerial report or whistleblowing (i.e., $r = \theta$ or $w = \theta$) will force the principal's beliefs to be consistent. Consistency may apply to the principal's final beliefs μ, as well as to her intermediate pre-whistleblowing beliefs μ_w and pre-report beliefs μ_r.

To fix ideas, it is helpful to use the example of a criminal investigation. A field officer's effort might lead to a high- or low-quality investigative lead (θ). An approval yields evidence in each period, while rejection yields nothing. An aggressive manager then might be motivated by the maximization of prosecutions, while a conservative manager might be more concerned about procedural protections. For the aggressive manager, the high-quality lead is desirable because it is aligned. But even if the lead is low-quality, the principal might have inconsistent beliefs if the evidence produced in period 1 were good.

In standard fashion, the discussion proceeds from the end of the game. To keep the notation for strategies manageable, I omit dependencies on information sets whenever the meaning of the expression in question is clear.

Period 2 Approval

Given any history of play $h_2 \in H_2$, the optimal approval strategy of the player i possessing period 2 decision-making rights is simply to approve any project yielding positive expected utility:

$$a_2^* = \begin{cases} 1 & \text{if } b^i E[x_2|h_2] - k^i \geq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

This expression implies that if she is retained ($s = M$), an aggressive manager will approve any project regardless of quality. Similarly, a conservative manager would reject the project regardless of quality. If the manager is fired ($s = P$) and neither the manager nor the employee reveal θ, then the principal's expected outcome is $E[x_2|h_2] = \mu\overline{x} + (1-\mu)\underline{x}$, where μ is her posterior belief that the project is of high quality. Otherwise, $E[x_2|h_2]$ will be $\underline{x}$ or $\overline{x}$ for the low- and high-quality projects, respectively. In their reporting and whistleblowing decisions, the manager and employee must therefore anticipate the politician's reaction to her updated knowledge of θ. These incentives—coupled with managerial punishments that induce performance by the employee—will in turn affect the employee's effort level.

Principal's Assignment of Decision Rights

Immediately preceding period 2, the principal chooses whether to override the manager's authority, which in turn depends on her posterior beliefs μ over the likeliness of high quality. These beliefs are consistent if either the manager or employee revealed θ, but may not be if they were silent. Using (5), the principal can determine whether her preferred action matches that of the manager's, and hence whether the manager should be retained. Since the principal breaks ties in favor of retaining managerial control, her optimal assignment of decision rights is:

$$s^* = \begin{cases} M & \text{if } x^M < \overline{x} \text{ and } \mu \geq \tilde{\mu} \\ & \quad \text{or } x^M > \underline{x} \text{ and } \mu \leq \tilde{\mu} \\ P & \text{if } x^M < \underline{x} \text{ and } \mu < \tilde{\mu} \\ & \quad \text{or } x^M > \overline{x} \text{ and } \mu > \tilde{\mu}. \end{cases} \tag{6}$$

As an example, in a criminal investigation, a field officer or her manager could inform the principal of the quality of a lead with certainty, or they may allow her to infer this information from the publicly available evidence. If the evidence is promising, or if a manager or whistleblower explicitly reveals that a case is of high quality, then the principal will be confident of a high-quality case. Knowing that an aggressive manager would continue to pursue this case, the principal retains her in period 2. Likewise, the principal will override a conservative manager, who would reject even good cases. The logic is reversed if the observable evidence is not promising, or if the informed players reveal a low quality case: the principal would retain the conservative manager and override an aggressive one.

Managerial Punishment

Because punishment is costless for the manager, any choice of punishment level p is optimal at the

https://doi.org/10.1017/S0003055408080192 Published online by Cambridge University Press

punishment stage. Clearly, however, she would like to use the punishment to induce optimal employee behavior. The manager therefore has two objectives in designing her punishment strategy. First, she wishes to affect the employee's effort level. Second, she may also wish to prevent whistleblowing.

To achieve the latter objective, the manager must threaten a punishment of at least the employee's expected gain from doing so. It will therefore be convenient to define the difference between the employee's expected payoff in a single period from an approved project of quality θ and zero (i.e., the payoff from outcome q):

$$l(\theta) = \left| b^E \int_X x f(x \mid \theta)\, dx - k^E \right|. \qquad \textbf{(7)}$$

By whistleblowing, the employee can potentially gain $\delta l(\theta) - c_w$ from a change in managerial authority. The manager can therefore *prevent* whistleblowing—essentially, issuing a "gag order"—by threatening an additional punishment of $\delta l(\theta) - c_w$ for doing so.

This combination of objectives yields two possible equilibrium punishment schemes. Consider initially the extreme case where the principal's capacity to punish is low; that is, $\overline{p} < \delta l(\theta) - c_w$. This is equivalent to an environment in which the employee cared greatly about policy outcomes. Since the manager would be unable to deter whistleblowing, she can condition the punishment only on some combination of θ and the first period policy outcome x_1. The optimal strategy is to ignore x_1 and focus exclusively on θ:

$$p^*_\theta(x_1, \theta, w) = \begin{cases} \overline{p} & \text{if } x^M < \underline{x} \text{ and } \theta = \underline{\theta}, \\ & \quad \text{or } x^M > \overline{x} \text{ and } \theta = \overline{\theta} \\ 0 & \text{otherwise.} \end{cases} \qquad \textbf{(8)}$$

I call this punishment strategy *quality-based punishment*. The manager punishes maximally for realizations of quality because this generates the greatest incentive for employee effort. Conditioning on x_1 is less effective because any value of x_1 could arise from either quality level. An aggressive manager punishes maximally for realizations of low quality because this generates the greatest incentive to maximize effort. In turn, this maximizes the chances of producing the aligned (high) quality level. Somewhat counterintuitively, a conservative manager punishes for *high* quality because now low quality is aligned. For example, a prosecutor might prefer good insider trading cases to bad ones, but may also be ideologically opposed to pursuing insider trading in general. She can discourage good cases by imposing excessive procedural cautions on investigators before seeking an indictment. In so doing, she raises the chances of a bad case, for which there would be more external support for rejection.

Now suppose that the manager's capacity for punishment is relatively high, so that $\overline{p} \geq \delta l(\theta) - c_w$. Though the manager may still punish based on quality, she may also deter "harmful" whistleblowing, which reveals inconsistent beliefs that would have allowed her to retain control in period 2. This strategy, termed *whistleblowing-based punishment*, entails threatening a punishment of $\delta l(\theta) - c_w$ for harmful whistleblowing. The remainder of the punishment capacity is "reserved" for punishing based on quality. Recalling that μ_w is the principal's belief of high quality immediately prior to the employee's whistleblowing decision, the strategy can be written formally as:

$$p^*_w(x_1, \theta, w) = \begin{cases} \overline{p} & \text{if } \theta \text{ unaligned and either } \mu_w \\ & \quad \text{consistent or } w = \theta \\ \overline{p} - \delta l(\theta) + c_w & \text{if } \theta \text{ unaligned, } \mu_w \text{ inconsistent} \\ & \quad \text{and } w = \emptyset \\ 0 & \text{otherwise.} \end{cases} \qquad \textbf{(9)}$$

Whistleblowing-based punishment draws attention to a subtle limitation on the manager's abilities. Even though suppressing a whistleblower might remove the manager's downside risk from the revelation of low quality, it introduces a problem for her allocation of retaliation resources. The manager cannot punish maximally for *both* quality and whistleblowing. Thus whistleblowing-based punishment requires the substitution of retaliation away from punishing low quality.

Whistleblowing

When the manager's punishment $p(x_1, \theta, \theta)$ does not exceed the net gain $\delta l(\theta) - c_w$ of revealing θ, whistleblowing allows the employee to reveal information that leads the principal to "fire" the manager and take over the project. The act of whistleblowing (i.e., choosing $w = \theta$) fully reveals project quality, and thus the employee will blow the whistle only if the principal's beliefs are both inconsistent and disadvantageous to the employee. Note that because a managerial report would have revealed θ, inconsistent beliefs could only result from the manager's silence.

To see the intuition of this decision, it is helpful to consider the circumstances under which the employee's strategy "separates" by choosing different actions for different quality levels. Suppose that a principal overseeing a law enforcement agency is optimistic of a high quality lead because the initial evidence is good (i.e., x_1 is high) and the aggressive manager who approved the investigation is silent. If the lead were in fact low quality, then the principal's beliefs would be inconsistent. Costs permitting, a field agent under the manager would have a clear incentive to blow the whistle.[12] If the lead were high quality, the field agent would gain

[12] Somewhat ironically, the employee's revelation is at least partially attributable to her own effort. It is possible that the employee could face consequences for this in a more elaborate model. Note, however, that even without whistleblowing, the principal conjectures correctly about the employee's effort and understands (with the exception of the corner case in which $e^* = 1$) that low quality occurs with positive probability. Additionally, in most public sector bureaucratic settings, a politically appointed manager is much more easily replaced than a civil servant.

https://doi.org/10.1017/S0003055408080192 Published online by Cambridge University Press

nothing by revealing θ. To the benefit of the field agent, the principal's beliefs become consistent: she not only directly learns when the lead is bad but also infers a good lead from the field agent's silence. By contrast, if the field agent remained silent regardless of quality, then the principal would have no basis for updating her beliefs about θ.

The employee will similarly wish to reveal that a conservative manager will cancel a high-quality project. More formally, the optimal whistleblowing strategy is then:[13]

$$w^* = \begin{cases} \theta & \text{if } p(x_1, \theta, \theta) < \delta I(\theta) - c_w, \text{ and either} \\ & \quad x^E > \underline{x}, \theta = \underline{\theta} \text{ and } \mu_w \text{ inconsistent; or} \\ & \quad x^E < \overline{x}, \theta = \overline{\theta} \text{ and } \mu_w \text{ inconsistent} \\ \emptyset & \text{otherwise.} \end{cases} \quad \textbf{(10)}$$

Managerial Report

The manager's reporting incentives are analogous with the employee's whistleblowing incentives. Generally, she will report θ instead of remaining silent whenever doing so will reverse the principal's inconsistent beliefs and allow her to retain managerial control. As with whistleblowing, the report fully reveals project quality and so the manager cannot lie.

The reporting strategy depends on whether the manager is aggressive or conservative. If she is aggressive, then she wishes to convince the principal that the project is of high quality, as both players would then agree to approving the period 2 project. Recall that μ_r is the principal's posterior belief of high quality prior to the manger's report r. Regardless of the true quality level, the manager does best by not reporting θ if the principal is optimistic of high quality ($\mu_r \geq \tilde{\mu}$). This causes the principal not to update her beliefs and to allow the manager to approve the project again in period 2, unless a whistleblower reveals a low-quality project. Thus an aggressive manager will report θ only when the quality is high and the principal's beliefs are inconsistent ($\mu_r < \tilde{\mu}$).

Symmetrically, a conservative manager wishes to convince the principal of low quality, as both players would then agree to allowing the manager to cancel the project in period 2. The manager's strategy therefore mirrors the aggressive case, with a report occurring only if principal is optimistic that project quality is high when it is in fact low. The optimal reporting strategy is then:

$$r^* = \begin{cases} \theta & \text{if } x^M < \underline{x}, \theta = \overline{\theta} \text{ and } \mu_r \text{ inconsistent; or} \\ & \quad x^M > \overline{x}, \theta = \underline{\theta} \text{ and } \mu_r \text{ inconsistent} \\ \emptyset & \text{otherwise.} \end{cases} \quad \textbf{(11)}$$

[13] This expression and the preceding discussion make use of the "minimum reporting" equilibrium refinement, which simplifies the analysis by selecting the equilibrium with no reporting or whistleblowing whenever possible.

The effect of this reporting strategy is to ensure that the principal is informed about the project quality whenever the quality level is aligned. However, the principal is "deceived" if the first period outcome x_1 yields inconsistent beliefs about an unaligned quality level. For example, a police manager interested in pursuing both good and bad investigative leads will happily allow the principal to draw her own conclusions on a lucky draw of evidence from a bad lead. Of course, this belief may be remedied at the subsequent whistleblowing stage.

It is evident that the employee's and manager's revelation incentives occasionally overlap. For instance, when the manager is aggressive and $x^E < \overline{x}$, both players wish to reveal a high-quality project when the principal's beliefs are inconsistent (i.e., $\theta = \overline{\theta}$ and $\mu_r < \tilde{\mu}$). Thus there exist equilibria in which either player reports θ. All such equilibria are identical in the amount of information reported; that is, whenever both players wish to reveal θ, one player will always do so. The "minimum reporting" equilibrium refinement selects the manager to report in these situations.[14]

Period 1 Approval

As in the second period, an aggressive manager has a myopic incentive to approve the project, whereas a conservative manager has a myopic incentive to reject it (analogously to (5)). However, in the first period, the manager could conceivably do the reverse in an effort to inform the principal about θ. This turns out not to be the case because the manager's subsequent reporting strategy always informs the principal when it is in the manager's interest to do so.

To illustrate why the manager would not approve or reject a project strategically, suppose that an aggressive manager adopts a "separating" strategy of approving only high-quality projects. This would ensure the consistency of the principal's beliefs, and thus the manager's authority would be revoked if and only if the quality were low. But the manager's reporting strategy (11) would inform the principal of high quality regardless of the approval strategy, whereas under the separating strategy the manager may not benefit from inconsistent beliefs when the quality is low. The manager therefore cannot do better than a myopic approval strategy:

$$a_1^* = \begin{cases} 1 & \text{if } b^M E[x_1|\theta] - k^M \geq 0 \\ 0 & \text{otherwise.} \end{cases} \quad \textbf{(12)}$$

[14] The "minimum reporting" refinement plays two additional roles. First, there are equilibria in which M reports θ when $\theta = \overline{\theta}$ for any subset of $\{x_1 \mid x_1 > \tilde{x}\}$ (see Proposition 1 for the characterization of $\tilde{x}$). Given that P expects truthful reporting for any such x_1, P would infer low quality from silence by M, and therefore M must report θ. It is clear that all equilibria in which M reports in this way are suboptimal for M. Additionally, all equilibria *except* the one in which M always reports θ when $\theta = \overline{\theta}$ are qualitatively similar to the one described here, in that M takes advantage of high realizations of x_1 to retain authority even when $\theta = \underline{\theta}$. Second, the refinement maintains consistency in managerial reporting strategies as c_w increases beyond the level at which whistleblowing is plausible.

https://doi.org/10.1017/S0003055408080192 Published online by Cambridge University Press

This strategy simplifies the equilibrium characterization because it is independent of θ for both aggressive and conservative managers. Because the approval decision is completely uninformative, the principal does not use it to update her beliefs. Thus her beliefs, prior to the managerial report r, are:

$$\mu_r = \begin{cases} e & \text{if } a_1 = 0 \\ \frac{f(x_1 \mid \bar{\theta})e}{f(x_1 \mid \bar{\theta})e + f(x_1 \mid \underline{\theta})(1-e)} & \text{if } a_1 = 1. \end{cases} \quad \textbf{(13)}$$

Employee Effort—Aggressive Manager

Generally, the employee must balance the cost of her effort against the possible outcomes induced by that effort. Obviously, effort directly affects θ. Since an aggressive manager approves any project in period 1, effort also indirectly affects outcomes through θ's effect on the initial policy outcome x_1.

Upon observing x_1, the principal forms a posterior belief of high quality, μ_r. By the MLRP property (2) and Bayes' Rule, μ_r is increasing in x_1. This implies the existence of a cutoff standard $\tilde{x}$ that the principal uses to assess the likelihood of each quality level.[15] When $x_1 < \tilde{x}$, then $\mu_r < \tilde{\mu}$ and the principal infers that quality is probably low. Similarly, higher realizations of x_1 are associated with high quality. Thus a key consideration in the employee's effort choice is its effect on having x_1 reach the standard $\tilde{x}$. In the example of the criminal investigation, a field agent who knows that whistleblowing is prohibitively costly might have an additional incentive to generate good leads, because she will not be able to inform the principal *ex post* of low quality.

The employee's effort choice also depends on the relative cost of whistleblowing, c_w. In the simplest case, if whistleblowing is sufficiently costly (i.e., $c_w > \delta l(\underline{\theta})$), then it is dominated by remaining silent. It will be useful to describe whistleblowing as *infeasible* in these situations, and *feasible* otherwise. Feasibility depends on the policy utility at stake, and might be enhanced by laws that protect or facilitate whistleblowing.

If whistleblowing is infeasible, then the manager has no need to deter whistleblowing and punishes based on quality. If whistleblowing is feasible, then the manager could also punish based on whistleblowing if she has sufficient capacity to deter it (i.e., $\overline{p} \geq \delta l(\underline{\theta}) - c_w$). Interestingly, these two punishment schemes induce the *same* effort level, since the employee effectively faces a penalty of $\overline{p}$ for a realization of low quality. But despite the commonality of effort, the punishment schemes yield very different outcomes. Whistleblowing may occur only when it is feasible and the manager punishes based on quality, and thus these are the only conditions under which the principal's beliefs are ultimately consistent and the "correct" approval decision is ensured in period 2. By contrast, if whistleblowing is infeasible or explicitly deterred, then the manager's reporting strategy allows inconsistent beliefs about a low-quality project. As a result, the principal takes the correct action whenever the project quality is high, but may allow the manager to retain authority when it is low.[16] Clearly, then, the principal would prefer that the manager punish based on quality.

The first main result formally ties the preceding observations together to show the existence of $\tilde{x}$ and characterize the equilibrium effort and punishment strategies. Loosely speaking, when whistleblowing is feasible the manager punishes based on whistleblowing if possible. Returning to the criminal investigation example, this means that the manager will generally deter a field agent from informing a politician of a "bad" case. The politician can then assess the manager's decision based only on his or her (necessarily biased) report and the public evidence generated indirectly by the field agent. Consequently, the politician may inadvertently allow an overzealous manager to pursue a weak case because of an apparently strong lead, when a whistleblower might have revealed damaging procedural irregularities in the evidence gathering process. The same informational problems arise when whistleblowing is infeasible, although the impossibility of whistleblower complaints allows the manager to focus exclusively on inducing higher effort from the field agent. Otherwise, if deterring whistleblowing is impossible, the manager punishes based on quality. This allows either the manager or the field agent to fill any gaps in the politician's knowledge about the quality of the case. Figure 1 illustrates the equilibrium under the two punishment strategies.

Proposition 1. *Principal's Standard, Employee Effort and Punishment Under an Aggressive Manager. There exists some $\tilde{x} \in X$ such that $\mu < \tilde{\mu}$ for $x_1 < \tilde{x}$ and $\mu > \tilde{\mu}$ for $x_1 > \tilde{x}$.*

E's effort is e^*

$$= \begin{cases} \frac{(1+\delta)b^E(\tilde{x}-\underline{x}) + \delta F(\tilde{x}|\underline{\theta})(b^E\underline{x} - k^E) + \overline{p}}{2c} & \text{if } c_w > \delta l(\underline{\theta}) \\ \frac{b^E(\tilde{x}-\underline{x}) + \delta(b^E\tilde{x} - k^E) + (1 - F(\tilde{x}|\underline{\theta}))c_w + \overline{p}}{2c} & \text{otherwise.} \end{cases}$$

If $\overline{p} < \delta l(\underline{\theta}) - c_w$, M's punishment is

$$p^*(x_1, \theta, w) = \begin{cases} \overline{p} & \text{if } \theta = \underline{\theta} \\ 0 & \text{if } \theta = \bar{\theta}, \end{cases} \quad \textit{and if } \overline{p} \geq \delta l(\underline{\theta}) - c_w,$$

$$p^*(x_1, \theta, w) = \begin{cases} \overline{p} & \text{if } \theta = \underline{\theta} \text{ and either } x_1 < \tilde{x} \text{ or } w = \theta \\ \overline{p} - \delta l(\underline{\theta}) + c_w & \text{if } \theta = \underline{\theta}, x_1 \geq \tilde{x}, \text{ and } w = \emptyset \\ 0 & \text{if } \theta = \bar{\theta}. \end{cases}$$

■

Proof. All proofs are in the Appendix. ■

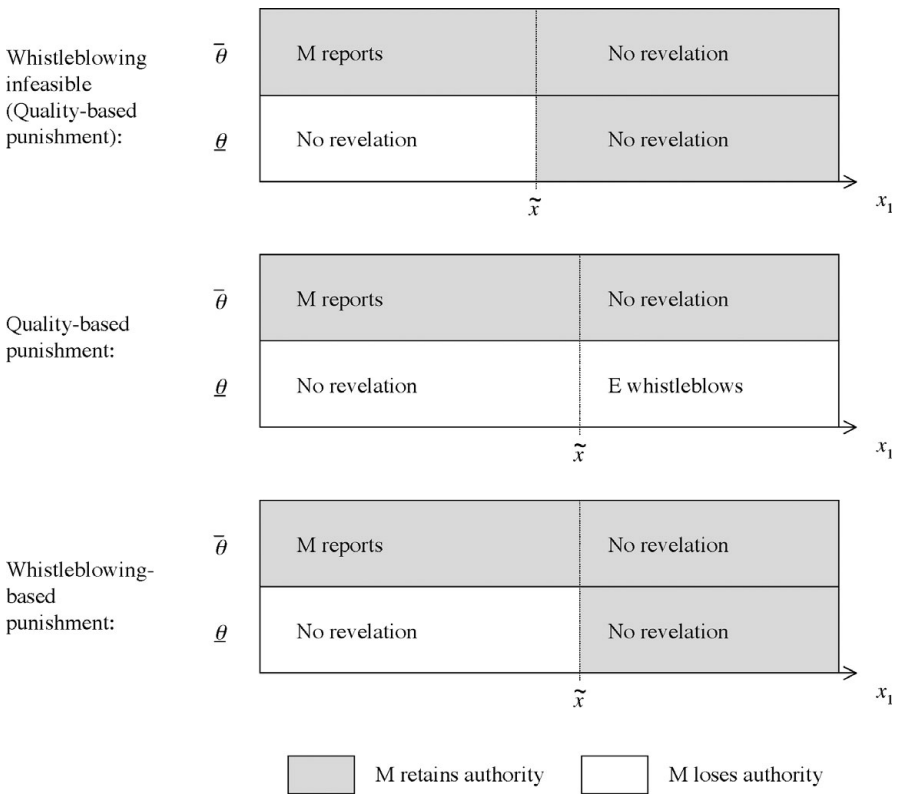
The expressions for e^* in Proposition 1 reveal two intuitive relationships. First, the effect of the punishment scheme is to raise effort, and effort is increasing in

[15] The model's assumptions do not rule out the possibility that $\tilde{x}$ is nonunique, and hence multiple equilibria may exist. For simplicity I select the lowest such value.

[16] If the strategy of punishing based on whistleblowing imposed a fixed cost on the manager, then she might prefer punishing based on quality when m is sufficiently low.

https://doi.org/10.1017/S0003055408080192 Published online by Cambridge University Press

FIGURE 1. Whistleblowing With an Aggressive Manager



Note: Under an aggressive manager, quality level $\bar{\theta}$ is aligned. When whistleblowing is infeasible, M punishes based on quality. P does not always learn of low quality, and allows M to retain control unless x_1 is low and M is silent. Otherwise, M can punish based on quality or whistleblowing. If M punishes based on quality, then E is free to whistleblow, and does so when x_1 is high and $\theta = \underline{\theta}$. If M punishes based on whistleblowing, then behavior resembles the infeasible-whistleblowing case and P may be deceived about θ. Note that $\tilde{x}$, the highest outcome at which P infers $\theta = \underline{\theta}$, is generally lower when whistleblowing is infeasible.

punishment capacity $\bar{p}$. This reflects the manager's desire to achieve high quality, as the principal and manager agree on approving only that kind of project. Second, a higher standard induces *lower* effort, because the employee prefers receivership to continued managerial control under low quality, and therefore faces a smaller downside risk when the cutoff is high.[17]

Somewhat less intuitively, the result reveals a central tension behind whistleblowing: effort is maximized when whistleblowing is infeasible. When whistleblowing is infeasible, the employee faces the full brunt of the manager's disciplining abilities, as well as the possible approval of a low-quality project in the second period. When whistleblowing is feasible, the employee's incentive to exert effort is reduced by one of two factors. If manager punishes based on quality, then the employee faces no risk of an approval of a low-quality project in the second period. If the manager can punish based on whistleblowing, then her disciplining authority is diverted away from inducing performance.

[17] It can also be shown that $\tilde{x}$ is decreasing in $\bar{p}$, and thus the principal is less permissive as the manager's disciplining ability grows.

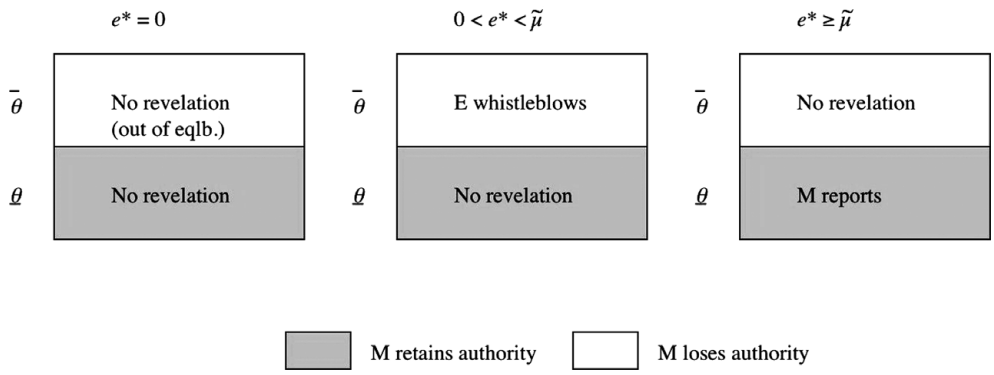
Employee Effort—Conservative Manager

As in the aggressive manager case, the employee balances the cost of effort and the expected payoff from each quality level. However, a conservative manager introduces two main differences to the players' strategies. First, because low quality is now aligned, the manager attempts to *discourage* high employee effort. This raises the probability of low quality, under which the principal would allow the manager to continue rejecting the project, instead of firing her and approving it. Second, because the manager now rejects any first period project, the principal loses an important piece of information present under an aggressive manager: there is no informative first period outcome x_1 that can be used to infer θ. In the criminal investigation example, the principal must guess the probable quality of a lead in the absence of actual evidence. Despite this, Bayes' Rule still allows the principal to infer initially that the probability of high quality is exactly the employee's effort level: $\mu_r = e^*$.

The equilibrium effort level depends primarily on this initial inference. If the employee's effort level is high so that $\mu_r > \tilde{\mu}$, then the principal is pessimistic of low quality and would revoke the manager's authority

https://doi.org/10.1017/S0003055408080192 Published online by Cambridge University Press

FIGURE 2. Whistleblowing With a Conservative Manager



Note: Under a conservative manager, quality level $\underline{\theta}$ is aligned. M will cancel all period 1 projects, and therefore P must infer θ from effort e. In equilibrium, P punishes based on quality. In the zero-effort corner case (far left), E does not have sufficient incentive to exert effort and whistleblow, and thus high quality cannot occur. Otherwise, either E or M will have an incentive to reveal θ, and thus P always learns θ.

unless she reports it. This may occur if the employee's intrinsic motivation to generate a high quality project, or b^E/k^E, is very high. Because the manager is willing report on low quality to correct the principal's inconsistent beliefs, the principal will be fully informed of project quality and the employee never needs to blow the whistle. The manager therefore punishes based on quality, penalizing high quality by $\overline{p}$.

Contrarily, if $\mu_r \leq \tilde{\mu}$, then the burden of proof is effectively reversed: now the employee has an incentive to whistleblow and reveal high quality. Thus if whistleblowing is feasible (i.e., $c_w \leq \delta l(\overline{\theta})$) or the manager punishes based on quality, then as in the previous case the principal will end up with consistent beliefs about θ. But if whistleblowing is either infeasible or explicitly suppressed, then the absence of an informative first-period outcome may force a "corner" effort of zero. To see why, suppose that the employee exerted a low but positive effort level, where $e < \tilde{\mu}$. This may occur if the employee's utility from an approved high-quality project, or equivalently b^E/k^E, is relatively low. The principal is then optimistic of low quality and inclined to continue the manager's control. Without whistleblowing (and now without the possibility of a low realization of x_1), there is no way for him or her to learn otherwise. Thus any effort would be wasted, and the employee does better with no effort.[18] The zero-effort solution is obviously the best outcome for a conservative manager: with no possibility of high quality, there is also no possibility of losing his or her authority.

The next result formalizes this argument to establish equilibrium punishment and effort strategies, which are illustrated in Figure 2. In the context of a criminal investigation, the principal's initial beliefs about the quality of a case would arise exclusively from her knowledge of the field agent's motivations. An office with a strong reputation for pursuing fraud cases might raise confidence about the quality of potential fraud cases but not narcotics cases. A principal would therefore be inclined toward overriding a manager's refusal to move forward with a fraud case, unless the manager reported her knowledge about the quality of the case. Similarly, he or she would be inclined toward supporting a manager's refusal to pursue a narcotics case in the absence of a whistleblower complaint to the contrary.

Proposition 2. *Effort and Punishment Under a Conservative Manager. E's effort is*:

$$e^* = \begin{cases} \frac{\delta(b^E\overline{x} - k^E) - \overline{p}}{2c} & \text{if } \frac{\delta(b^E\overline{x} - k^E) - \overline{p}}{2c} > \tilde{\mu} \\ \max\left\{0, \frac{\delta(b^E\overline{x} - k^E) - c_w - \overline{p}}{2c}\right\} & \text{otherwise,} \end{cases}$$

and M's punishment is

$$p^*(x_1, \theta, w) = \begin{cases} \overline{p} & \text{if } \theta = \overline{\theta} \\ 0 & \text{if } \theta = \underline{\theta}. \end{cases}$$ ■

Two differences with Proposition 1 are worthy of note. First, a conservative manager's punishment shifts effort downwards (in noncorner cases), and higher punishment capacity induces weakly lower effort. Second, whistleblowing-based punishment is unnecessary. This is a consequence of the lower bound on employee effort: the ability to punish based on whistleblowing also implies the ability to drive the employee's effort level down to the zero-effort corner case through simple quality-based punishment. The corner case and the lack of a first-period outcome yield much simpler equilibrium strategies.

Propositions 1 and 2 generate roughly opposite comparative statics relations between the cost of whistleblowing and effort. Under a conservative manager, effort is weakly decreasing in whistleblowing cost. This happens because whistleblowing allows the employee to reveal high quality even if her effort level is low, thus avoiding the corner solution and "wasted effort."

[18] Unlike the aggressive manager case, the feasibility of whistleblowing does not automatically change employee effort, since the corner solution may obtain even when whistleblowing is feasible.

https://doi.org/10.1017/S0003055408080192 Published online by Cambridge University Press

Low whistleblowing costs therefore increase the returns to employee effort. The two results therefore establish a simple relationship: *feasible whistleblowing moves effort in the opposite direction from that which the manager would prefer*.

The aggressive and conservative manager scenarios flesh out the central organizational incentive problems produced by whistleblowing. From the principal's perspective, whistleblowing may drive a wedge between *ex ante* and *ex post* incentives. Relative to a world in which whistleblowing is infeasible, the politician benefits from whistleblowing *given* a certain level of employee effort. From an *ex post* perspective, whistleblowing may have the salutary effect of causing managerial authority to be revoked when the manager would choose the "wrong" action in period 2. However, the *ex ante* effects on employee effort may not benefit the principal. The next section examines several implications of these results for whistleblower policy.

It is finally worth observing that while the allocation of employee incentives is of central interest in any discussion of organizational performance, the model excludes from consideration several plausible kinds of managerial behavior. For example, whistleblowing neither "disciplines" the manager to report θ more often, nor changes her (sincere) approval decisions. These do not occur in part because reports fully reveal θ, and the manager's preferences are common knowledge. Whistleblowing also cannot induce the manager to work harder at reaching a "good" decision. Relaxing the informational assumptions, or extending moral hazard problems to the manager, are logical extensions of the basic framework.

WHISTLEBLOWER POLICY

Personal Costs and Benefits of Whistleblowing

The preceding results are useful for deriving implications for whistleblower protection. Perhaps the most straightforward way to assess the impact of such protections is by examining the relationship between the principal's expected utility and the cost of whistleblowing, c_w. This approach is sensible when a principal may directly manipulate the hurdles that employees face in order to reveal information.

Legislatures have historically played a central role in determining these costs. Laws such as the WPA typically provide procedural guarantees such as promises of confidentiality to facilitate employee reporting.[19] Another centerpiece of American whistleblowing legislation is the FCA, which allows *qui tam* relators to receive a portion of the revealed fraud or damages. It is thought that this provides an important incentive for employees to come forward. (The law may also allow politicians to reclaim damages more easily, though this aspect will not be addressed here.) Thus, c_w might reflect both the personal cost associated with initiating a whistleblower complaint, and the expected monetary reward for pursuing a complaint successfully.[20]

The next result uses Propositions 1 and 2 to show that, perhaps surprisingly, the principal does not always benefit as whistleblowing becomes easier. Reducing c_w unambiguously helps the principal when the manager is conservative, but not when she is aggressive.

Comment 1. *Relative Cost of Whistleblowing.* (*i*) *If M is aggressive, there exists a* $\delta^* \geq 0$ *such that if* $\delta \leq \delta^*$, *P's expected utility is maximized by any* c_w *such that* $c_w > \delta l(\underline{\theta})$.

(*ii*) *If M is aggressive,* $\delta l(\underline{\theta}) > \overline{p}$ *and* $\tilde{x} \in \text{int}X$ *at* $c_w = \delta l(\underline{\theta}) - \overline{p}$, *then P's expected utility is non-increasing in the neighborhood of* $c_w = \delta l(\underline{\theta}) - \overline{p}$.

(*iii*) *If M is conservative, P's expected utility is weakly decreasing in* c_w. ■

Parts (i) and (ii) of the result reveal how laws such as the FCA generate *ex post* incentives for information revelation that may conflict with optimal *ex ante* incentives for performance. The conflict occurs under an aggressive manager. Part (i) shows that when the first-period outcome matters relatively more (though δ^* may be infinite under some conditions), the principal effectively wishes to ban whistleblowing by making it infeasible. This allows the manager to "front load" the employee's incentives to produce a high-quality project to the maximum extent possible, which in turn maximizes the chances of an initial policy success.

Part (ii) shows that the principal does not unambiguously prefer high whistleblowing costs. The intuition of the result is that the manager punishes based on whistleblowing if possible, as this allows her to take advantage of favorable realizations of x_1. At $c_w = \delta l(\underline{\theta}) - \overline{p}$, the manager is barely able to punish based on whistleblowing. Thus the principal would prefer to have c_w below this threshold (which makes whistleblowing possible) than above. In other words, in this region of c_w, the principal's *ex post* concern for information revelation dominates. The result requires only the modest condition that the principal's equilibrium performance standard $\tilde{x}$ not be trivial in this region, in the sense that her beliefs about quality would not be affected by the first-period outcome.

As part (iii) establishes, when the manager is conservative, there is no tension between *ex post* and *ex ante* incentives. In this case managerial "wrongdoing" occurs only for a high-quality project, which is associated with both higher effort and whistleblowing. For the employee, high whistleblowing costs and high effort work against one another: unlike the aggressive manager case, high effort raises the likelihood of having to pay c_w. Thus lower values of c_w unambiguously help both the principal and the employee.

[19] The WPA succeeded the Civil Service Reform Act of 1978, which was one of the first whistleblower protection laws of its kind. Prior to the passage of these laws, whistleblower protections were typically handled by courts.

[20] The two interpretations of c_w raise an important subtlety. Procedural protections affect all whistleblower cases, while *qui tam* provisions affect only successful ones. In the equilibrium of this model, these coincide, but in practice (or in a more elaborate model) they may not.

https://doi.org/10.1017/S0003055408080192 Published online by Cambridge University Press

TABLE 1. Examples of Managerial Strategies

Whistleblowing is:	Feasible			Infeasible	
$\overline{p}$	0.16	0.2		0.16	0.2
Punishment Strategy	Quality	Quality*	Whistle.	Quality	Quality
Cutoff Standard	—	—	0.699	0.664	0.580
Effort	0.243	0.263	0.263	0.274	0.301
M Expected Payoff	0.841	0.869	0.968	0.993	1.054
P Expected Payoff	0.017	0.026	0.006	0.009	0.018

*Out of equilibrium.

Comment 1 yields several implications for institutional design. A principal would want to minimize the personal cost of blowing the whistle in a "conservative" agency. Possible institutional measures include compensation for damages, redundant venues for pursuing complaints, and liability protection. By contrast, she may desire much higher procedural hurdles in an "aggressive" agency. This can occur when δ is low, which corresponds to a project that is relatively uncommon or idiosyncratic. In this environment, the principal values a successful initial outcome highly and would do better if managers did not have to worry about whistleblowers. This logic is consistent with the weaker protections observed in U.S. national security agencies.[21] However, banning whistleblowing outright is not always optimal, as the principal may in other circumstances want to ensure that whistleblowing is not deterred.

By tying the cost of whistleblowing with the preferences of managers in the bureaucracy, the result may also help to explain some of the historical variation in *qui tam* laws. The percentage of damages which *qui tam* relators could claim has varied considerably over history (e.g., Park 1991). The original 1863 False Claims Act allowed up to 50%. Amendments in 1943 addressed what Congress believed to be "parasitic" whistleblowers by decreasing the relator's share to 10%–25%, depending on the type of case. The law was reinvigorated by 1986 amendments that raised the share to 15%–30%. Along with new WPA provisions which allowed plaintiffs to take some cases directly to court, this reform increased recovered amounts from less than $10 million to over $100 million per year.

Numerical Example. To illustrate this result, and in particular how whistleblowing protections might lower the politician's expected payoffs, suppose that M is aggressive, with $b^M = 3$, $b^P = 1.8$, $b^E = 1.6$, and $k^M = k^E = k^P = 1$. Let $\delta = 0.9$, $c = 1$, and $c_w = 0$ when whistleblowing is feasible. The policy space is $X = [0, 1]$. The outcomes are distributed uniformly for quality $\underline{\theta}$; $f(x_t \mid \underline{\theta}) = 1$, whereas the density is linear for quality $\overline{\theta}$; $f(x_t \mid \overline{\theta}) = 2x_t$. It is straightforward to calculate that $\underline{x} = 0.5$, $\overline{x} = 0.667$, and $\tilde{\mu} = 0.333$.

[21] See Congressional Research Service report RL33215 (2005) for an overview.

Table 1 compares effort and payoffs for feasible and infeasible whistleblowing. It also varies $\overline{p}$ between 0.16, which forces M to punish based on quality, and 0.2, which allows M to punish based on whistleblowing (i.e., $0.16 < \delta l(\underline{\theta}) < 0.2$).

When $\overline{p} = 0.2$, M's expected payoff is higher when she deters whistleblowing, even though effort is identical under both punishment strategies. This is because the cutoff standard for inferring that $\mu = \tilde{\mu}$ is $\tilde{x} = 0.699$, which gives M an over 30% chance of retaining managerial control even when the project is low quality. P would prefer quality-based punishment, since this would effectively always reveal θ in period 2. Given that M does not do this, P would be better off if whistleblowing were infeasible. The impossibility of whistleblowing does not reduce information revelation and raises effort because managerial resources are not diverted toward deterring whistleblowing.

When $\overline{p}$ is reduced to 0.16, M's more limited ability to punish reduces effort. Because punishment can only be based on quality, E blows the whistle whenever P has inconsistent beliefs. P therefore always ends up with consistent beliefs in period 2. As Comment 1 predicts, effort is lower when whistleblowing is feasible because E faces a lower downside from a low quality project. Nevertheless, the informational gains from whistleblowing more than offset the lost effort. Thus the ability to whistleblow helps the politician.

Optimal Managerial Punishments

In addition to the costs borne by individual employees, whistleblower protection laws also address the ability of managers to punish them. Laws such as the WPA contain provisions such as injunctions against managerial actions that make retaliation against whistleblowers more difficult, if not illegal. Section 8547.3(a)–(c) of the California Whistleblower Protection Act provides one example:

> Use or attempted use of official authority or influence to interfere with disclosure of information; prohibition; civil liability
> (a) An employee may not directly or indirectly use or attempt to use the official authority or influence of the employee for the purpose of intimidating, threatening, coercing, commanding, or attempting to intimidate, threaten, coerce, or command any person for the purpose of interfering with the rights conferred pursuant to this article.

https://doi.org/10.1017/S0003055408080192 Published online by Cambridge University Press

(b) For the purpose of subdivision (a), "use of official authority or influence" includes promising to confer, or conferring, any benefit; effecting, or threatening to effect, any reprisal; or taking, or directing others to take, or recommending, processing, or approving, any personnel action, including, but not limited to, appointment, promotion, transfer, assignment, performance evaluation, suspension, or other disciplinary action.
(c) Any employee who violates subdivision (a) may be liable in an action for civil damages brought against the employee by the offended party.

Given that some level of managerial discretion is inevitable in practice, it is useful to ask what limitations on managerial actions principals would favor. One desideratum might be that of discouraging managers from punishing based on whistleblowing. As the analysis of the previous section established, a switch to punishing based on quality will benefit the principal. Thus, politicians would stand to benefit if whistleblowing laws have the effect of regulating *only* the kinds of strategies employed by managers. Unfortunately, the model developed here also suggests that this might be difficult to accomplish. In addition to obstructing retaliation against an employee, whistleblower protection laws can also change the employee's incentive to blow the whistle. In particular, an employee could invoke whistleblower protections to reduce the scope of *all* managerial disciplinary action.[22]

I formalize this idea with a simple scenario, where an employee can automatically invoke whistleblowing protections simply by choosing $w = \theta$. This might correspond to an environment in which even weak complaints take a long time to resolve, thus forestalling potentially legitimate managerial actions. Suppose that the maximum legal punishment against a whistleblower were reduced to $\overline{p}^w < \overline{p}$, while leaving the maximum punishment for nonwhistleblowers at $\overline{p}$. If $\overline{p}^w < \delta l(\theta) - c_w$, then the manager is prevented from punishing based on whistleblowing. But the reduction imposes another problem on the manager: if $\overline{p} > \overline{p}^w + c_w$, then the employee can circumvent a high punishment simply by becoming a whistleblower. In other words, when $\overline{p}^w$ is sufficiently low the manager would be unable to deliver the full punishment of $\overline{p}$ even when the employee would normally not blow the whistle. As a result, punishments cannot effectively exceed $\overline{p}^w$ in an optimal punishment strategy.

Comment 2. *Whistleblower Protection and Managerial Latitude. If $\overline{p}^w + c_w < \overline{p}$, then in the aggressive and conservative manager cases there exists an optimal punishment strategy satisfying $p^*(x_1, \theta, w) \leq \overline{p}^w + c_w$ for all x_1, θ, and w.* ■

Thus, unless $\overline{p}^w$ is "close" to $\overline{p}$, limiting whistleblower-only retaliation to $\overline{p}^w$ effectively constrains *all* punishments to be no greater than $\overline{p}^w$. Any such limitation is then effectively a reduction in $\overline{p}$. Analogously to Comment 1, the next comment characterizes the effect of $\overline{p}$ (and equivalently, low values of $\overline{p}^w$) on the principal's utility: reduced managerial latitude may harm the principal under an ambitious manager, but help under a conservative manager.[23]

Comment 3. *Optimal Managerial Latitude.* (*i*) *If M is aggressive, P's expected utility is increasing in $\overline{p}$ for $\overline{p} < \delta l(\underline{\theta}) - c_w$.*

(*ii*) *If M is aggressive, $\delta l(\underline{\theta}) > c_w$ and $\tilde{x} \in \text{int}X$ at $\overline{p} = \delta l(\underline{\theta}) - c_w$, then P's expected utility is non-increasing in the neighborhood of $\overline{p} = \delta l(\underline{\theta}) - c_w$.*

(*iii*) *If M is conservative, P's expected utility is weakly decreasing in $\overline{p}$.* ■

Parts (i) and (ii) of Comment 3 again illustrate the tension between the *ex ante* desire for performance and the *ex post* desire for information revelation under an aggressive manager. The principal is happy to give an aggressive manager more discretion for inducing high quality, as long as the manager is restricted to punishing based on quality. However, because the manager punishes based on whistleblowing whenever possible, her ability to do so will decrease the principal's utility. As a result, in the neighborhood of $\delta l(\underline{\theta}) - c_w$, the increase in effort from a higher value of $\overline{p}$ is outweighed by the inability to have every low quality project revealed (with the help of whistleblowing). By contrast, under a conservative manager, reducing managerial discretion simply reduces her ability to reduce employee effort. Thus restrictions on managerial actions are always desirable from the principal's perspective.[24]

A common feature of all modern civil service systems is their limitation on managerial discretion over employee payoffs. When whistleblowing is relatively costless, employees can use whistleblower protections to limit further the extent of managerial retribution. Whistleblower laws then essentially become de facto extensions of basic civil service protections. The consequences for employee incentives then depend on the preferences of the manager relative to the status quo.

CONCLUSIONS

The extent and effectiveness of whistleblower policies remain open questions. It is therefore worth remarking

[22] See Denise Kersten Wills, "You're Fired," *Government Executive* 38(3), March 1, 2006. One fact possibly consistent with the view that whistleblower laws can be invoked too frequently is that between fiscal years 1997 and 1999, only 16%–26% of cases given full review by the Office of Special Counsel resulted in favorable judgments (U.S. Office of Special Counsel 1999).

[23] An identical, but more cumbersome result can be proved for small changes in $\overline{p}^w$, such that $\overline{p}^w + c_w > \overline{p}$.

[24] In addition to having preferences over $\overline{p}$, P may have induced preferences over the manager's utility from office-holding, m. If the suppression of whistleblowing were costly, then an aggressive manager would punish based on quality when m is low (i.e., M is less career-minded or more policy-minded), and would punish based on whistleblowing if feasible when m is high (see footnote 16). The principal therefore prefers low-m managers, whose incentives are less distorted toward the preservation of managerial prerogatives. A high-m manager might correspond to a career civil servant, while low-m manager might correspond to a political appointee. One empirical implication is that strengthening whistleblower protections in an agency will increase the proportion of political appointees.

https://doi.org/10.1017/S0003055408080192 Published online by Cambridge University Press

on how the main variables of the model can be measured empirically. In addition to changes in the strength of federal laws, variations in whistleblower protections can be captured through contemporaneous variations in U.S. state laws. Variations in managerial preferences can be measured through the composition of agency personnel relative to that of their political principals (e.g., Lewis 2008). Highly politicized agencies that receive an influx of new funding and programs might be considered aggressive, while those that do not might be considered conservative. Managerial aggressiveness may also be linked to organizational structure. In the model, an aggressive manager is one who places more emphasis on avoiding Type II error (relative to the "default" policy of q), while a conservative manager is more concerned with avoiding Type I error. Thus there is ample room for exploiting the extensive literature on the impact of organizational design on Type I and II errors (Bendor 1985; Heimann 1997; Carpenter and Ting 2007).

For well over a century, observers of all kinds have recognized the importance of whistleblowers in aiding the transmission of information from bureaucracies. In the American public sector, this recognition reflects in part the ideal of neutrally competent bureaucrats who may play a role in mitigating the excessive politicization of bureaucracies. Good whistleblowing policy is thought to improve the monitoring of agencies, as well as to provide proper incentives to employees. The model helps to assess such policies by capturing many of the incentives faced by employees who might wish to reveal policy-relevant information, but face the prospect of reprisals from their immediate superiors. Perhaps as importantly, it also illustrates how internal organizational structure can be relevant to the political control of bureaucracies.

The model illuminates a central tension in the design of whistleblowing policy: the politician's *ex ante* desire for greater effort versus her *ex post* desire for information revelation. Generally speaking, the results suggest that whistleblower protections do very well on the latter, but relatively poorly on the former. The key variable is the relation between the preferences of the manager and those of the politician. An aggressive manager, who is more inclined than the politician to approve a project, will discipline employees in ways that increase their effort. In this case, whistleblower protections can reduce the power of employee incentives. In some circumstance it may even be optimal for a politician to disallow whistleblowing. As more information does not necessarily improve the principal's control, this conclusion is counterintuitive from the perspective of most existing theories of control of the bureaucracy. By contrast, conservative managers wish to suppress effort, and so whistleblower protections will have the salutary effect of increasing employee effort.

APPENDIX

The following lemmas formally establish parts of the equilibrium strategies that are necessary for deriving Propositions 1–2. The principal's optimal assignment of decision rights s^* (6) is established in the text and is omitted here.

Lemma 1. *Reporting and Whistleblowing.*

$$r^* = \begin{cases} \theta & \text{if } x^M < \underline{x}, \theta = \overline{\theta} \text{ and } \mu_r < \tilde{\mu}; \text{or} \\ & \quad x^M > \overline{x}, \theta = \underline{\theta} \text{ and } \mu_r > \tilde{\mu} \quad \textit{and} \\ \emptyset & \text{otherwise.} \end{cases}$$

$$w^* = \begin{cases} \theta & \text{if } p(x_1, \theta, \theta) < \delta l(\theta) - c_w, \text{ and either} \\ & \quad x^E > \underline{x}, \theta = \underline{\theta} \text{ and } \mu_w \geq \tilde{\mu}; \text{or} \\ & \quad x^E < \overline{x}, \theta = \overline{\theta} \text{ and } \mu_w \leq \tilde{\mu} \\ \emptyset & \text{otherwise.} \end{cases}$$ ■

Proof. Consider first the whistleblowing decision, w. Clearly, if $p(x_1, \theta, \theta) \geq \delta l(\theta) - c_w$, then $w^* = \emptyset$. If $p(x_1, \theta, \theta) < \delta l(\theta) - c_w$, then there are three subcases. First, if $x^E > \overline{x}$ (which implies M is aggressive), then E's period 2 payoff is maximized by $a_2 = 0$. By (5) and (6), $a_2^* = 0$ iff $\mu < \tilde{\mu}$. If $\mu_w < \tilde{\mu}$, then by the minimum reporting equilibrium selection rule, E chooses $w^* = \emptyset$ for all θ, which ensures that $\mu < \tilde{\mu}$ for all θ. If $\mu_w \geq \tilde{\mu}$, then $w = \emptyset$ for all θ cannot be optimal since it implies $\mu = \mu_w$ even if $\theta = \underline{\theta}$. Also, $w = \theta$ if $\theta = \overline{\theta}$ is obviously dominated, and so $w^* = \emptyset$ if $\theta = \overline{\theta}$. Thus E chooses $w^* = \theta$ when $\theta = \underline{\theta}$. Under this strategy, Bayes' rule implies that $\mu = 0 < \tilde{\mu}$ when $w = \theta$, which results in $a_2^* = 0$, and $\mu = 1 > \tilde{\mu}$ when $w = \emptyset$. Putting the two parts together, $w^* = \theta$ iff $\theta = \underline{\theta}$ and $\mu_w \geq \tilde{\mu}$. Note that this strategy implies that $w^* = \emptyset$ if $r^* = \theta$.

Second, if $x^E < \underline{x}$, a symmetric analysis establishes that $w^* = \theta$ iff $\theta = \overline{\theta}$ and $\mu_w \leq \tilde{\mu}$. Third, $x^E \in (\underline{x}, \overline{x})$, then the arguments of first two cases can be combined straightforwardly to show that $w^* = \theta$ if $\theta = \underline{\theta}$ $(= \overline{\theta})$ and $\mu_w > \tilde{\mu}$ $(< \tilde{\mu})$, and if $\theta = \underline{\theta}$ $(= \overline{\theta})$ and $\mu_w = \tilde{\mu}$ for M aggressive (conservative).

For the reporting decision r, note that M's payoff is maximized by $s = M$, which yields a period 2 payoff of at least $m > 0$. There are two subcases, depending on the location of x^M. First, if $x^M < \underline{x}$, then (6) implies that $s^* = M$ iff $\mu \geq \tilde{\mu}$. Suppose that $\mu_r < \tilde{\mu}$. If $\theta = \overline{\theta}$, then clearly under any equilibrium reporting and whistleblowing strategies, either $r^* = \theta$ or $w^* = \theta$ (resulting in $\mu = 1 \geq \tilde{\mu}$). The minimum reporting rule is therefore uniquely satisfied by: $r^* = \theta$ iff $\theta = \overline{\theta}$, which by Bayes' results in beliefs $\mu = 1$ (0) if $r = \theta$ $(= \emptyset)$. To verify that this is an equilibrium strategy, note that if $\theta = \overline{\theta}$ then M achieves her maximal payoff by $r = \theta$. If $\theta = \underline{\theta}$, then $r = \theta$ results in $\mu = 0$, and hence $s = P$ and $a_2 = 0$, which yields M's minimal period 2 payoff of 0. Now suppose $\mu_r \geq \tilde{\mu}$. Now the minimum reporting rule is uniquely satisfied by $r^* = \emptyset$ for all θ. To verify that this is an equilibrium strategy, note that if $\theta = \overline{\theta}$, then regardless of w^* the report cannot change s. If $\theta = \underline{\theta}$ then $r = \theta$ results in $\mu = 0$ and $s = P$. Thus $r^* = \theta$ iff $\theta = \overline{\theta}$ and $\mu_r < \tilde{\mu}$ is the unique reporting strategy satisfying minimum reporting.

Second, if $x^M > \overline{x}$, the result follows by symmetry with the first case. ■

Lemma 2. *Managerial Approval. For M in period 1 and player $i \in \{M, P\}$ with decision authority at period* 2,

$$a_1^* = \begin{cases} 1 & \text{if } b^M E[x_1|\theta] - k^M \geq 0 \\ 0 & \text{otherwise,} \quad \textit{and} \end{cases}$$

$$a_2^* = \begin{cases} 1 & \text{if } b^i E[x_2|h_2] - k^i \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$ ■

https://doi.org/10.1017/S0003055408080192 Published online by Cambridge University Press

Proof. At $t = 2$, the result follows straightforwardly from the discussion in Section 3. I therefore restrict attention to $t = 1$.

Suppose $x^M < \underline{x}$, so that M would myopically approve the project for any θ. If $\theta = \overline{\theta}$ and $a_1 = 1$, then according to M's reporting strategy (Lemma 1), $\mu \geq \tilde{\mu}$. Thus, by (6), P chooses $s^* = M$, and $a_2^* = 1$. M then receives $(1 + \delta)(b^M\overline{x} - k^M + m)$ by choosing $a_1 = 1$. By choosing $a_1 = 0$, M could expect at most $\delta(b^M\overline{x} - k^M) + (1 + \delta)m$, which is strictly worse. Thus $a_1^*(\overline{\theta}) = 1$ in any equilibrium.

Now consider whether M could choose $a_1(\underline{\theta}) = 0$. Given that $a_1^*(\overline{\theta}) = 1$, Bayes' Rule implies $\mu = 0$ if $a_1 = 0$ in any equilibrium, and so by (6) P chooses $s^* = P$ and $a_2^* = 0$. This results in a payoff of m for M. By deviating to $a_1 = 1$, M ensures herself a payoff of at least $b^M\underline{x} - k^M + m$. Thus $a_1^*(\underline{\theta}) = 1$ in any equilibrium.

The case where $x^M > \overline{x}$ is proved identically and is omitted. ■

Lemma 3. *Managerial Punishment. The optimal punishment strategy that conditions on θ and x_1 (quality-based punishment) is*

$$p_\theta^*(x_1, \theta, w) = \begin{cases} \overline{p} & \text{if } x^M < \underline{x} \text{ and } \theta = \underline{\theta}, \\ & \quad \text{or } x^M > \overline{x} \text{ and } \theta = \overline{\theta} \\ 0 & \text{otherwise.} \end{cases} \tag{14}$$

The optimal strategy that conditions on θ, x_1, and w (whistleblowing-based punishment) is

$$p_w^*(x_1, \theta, w) = \begin{cases} \overline{p} & \text{if } x^M < \underline{x}, \theta = \underline{\theta} \text{ and either} \\ & \quad \mu_w < \tilde{\mu} \text{ or } w = \theta; \\ & \quad \text{or } x^M > \overline{x}, \theta = \overline{\theta} \text{ and either} \\ & \quad \mu_w > \tilde{\mu} \text{ or } w = \theta \\ \overline{p} - \delta l(\underline{\theta}) + c_w & \text{if } x^M < \underline{x}, \theta = \underline{\theta}, \mu_w \geq \tilde{\mu} \\ & \quad \text{and } w = \emptyset \\ \overline{p} - \delta l(\overline{\theta}) + c_w & \text{if } x^M > \overline{x}, \theta = \overline{\theta}, \mu_w \leq \tilde{\mu} \\ & \quad \text{and } w = \emptyset \\ 0 & \text{if } x^M < \underline{x} \text{ and } \theta = \overline{\theta}; \\ & \quad \text{or } x^M > \overline{x} \text{ and } \theta = \underline{\theta}. \end{cases} \tag{15}$$

■

Proof. It is first necessary to determine which incentives M wishes to provide. Observe that using Lemmas 1, 2, and (6), it is possible given any punishment strategy $p(\cdot)$ to bound each player's expected utility conditional upon θ. Let $g^i(\theta, e)$ denote player i's expected payoff under optimal strategies given θ and conjectured effort e, excluding any punishments and E's effort cost. M's expected utility can be written in the following general form:

$$U^M = eg^M(\overline{\theta}, e) + (1-e)g^M(\underline{\theta}, e). \tag{16}$$

It is clear by (16) that M must provide incentives for higher (lower) effort if $\min_e g^M(\overline{\theta}, e) \geq (\leq) \max_e g^M(\underline{\theta}, e)$.

There are two cases, depending on the location of x^M.

(i) $x^M < \underline{x}$. By Lemma 2, $a_1^* = 1$. By (6) and Lemma 1, $\mu \geq \tilde{\mu}$ and $s^* = M$ whenever $\theta = \overline{\theta}$. Thus $g^M(\overline{\theta}, e) = (1+\delta)[b^M\overline{x} - k^M + 2m]$. If $\theta = \underline{\theta}$, then $g^M(\underline{\theta}, e) \leq (1+\delta)[b^M\underline{x} - k^M + 2m]$. Hence $\min_e g^M(\overline{\theta}, e) > \max_e g^M(\underline{\theta}, e)$, so M prefers higher effort.

First I establish the optimal punishment strategy that conditions on θ and x_1. Consider the set of integrable functions $\{p(x_1, \theta, w)\}$, where $p(x_1, \theta, w) \in [0, \overline{p}]$ for all x_1, θ, and w. Letting e^* denote the equilibrium effort level, E's objective is then:

$$\begin{aligned} U^E = {} & eg^E(\overline{\theta}, e^*) + (1-e)g^E(\underline{\theta}, e^*) \\ & - e\int_X p(x, \overline{\theta}, w)f(x \mid \overline{\theta})\,dx - (1-e) \\ & \times \int_X p(x, \underline{\theta}, w)f(x \mid \underline{\theta})\,dx - ce^2. \end{aligned} \tag{17}$$

This objective is concave in e. Clearly, effort is maximized by choosing $p(x_1, \theta, w)$ to maximize $\frac{dU^E}{de}$. Differentiating (17) reveals that $\frac{dU^E}{de}$ is maximized with respect to $p(x_1, \theta, w)$ when $\int_X p(x, \underline{\theta}, w)f(x \mid \underline{\theta}) - p(x, \overline{\theta}, w)f(x \mid \overline{\theta})\,dx$ is maximized. Hence an optimal strategy for quality-based punishment is $p(x_1, \underline{\theta}, w) = \overline{p}$ and $p(x_1, \overline{\theta}, w) = 0$ for all x_1 and w.

Next I establish the optimal punishment strategy that also conditions on w. By Lemma 1, $w = \theta$ may be an equilibrium strategy only if $\theta = \underline{\theta}$ and $\mu_w \geq \tilde{\mu}$. Thus conditioning on w can benefit M only if $\theta = \underline{\theta}$ and $\mu_w \geq \tilde{\mu}$. Because whistleblowing yields E an expected utility change of $\delta l(\underline{\theta}) - c_w$, the punishment must satisfy $p(x_1, \underline{\theta}, \theta) - p(x_1, \underline{\theta}, \emptyset) \geq \delta l(\underline{\theta}) - c_w$ to deter whistleblowing. Clearly, $p(x_1, \underline{\theta}, \theta) - p(x_1, \underline{\theta}, \emptyset) > \delta l(\underline{\theta}) - c_w$ is unnecessary to deter whistleblowing, and thus the optimal punishment conditioning on w must satisfy $p(x_1, \underline{\theta}, \theta) - p(x_1, \underline{\theta}, \emptyset) = \delta l(\underline{\theta}) - c_w$ when $\mu_w \geq \tilde{\mu}$.

Applying the above argument for conditioning on θ and x_1, M also punishes all realizations of $\theta = \overline{\theta}$ by zero, and $\theta = \underline{\theta}$ by the maximum possible amount. Thus if $\theta = \underline{\theta}$ and $\mu_w \geq \tilde{\mu}$, we have $p(x_1, \underline{\theta}, \emptyset) = \overline{p} - \delta l(\underline{\theta}) + c_w$, and if $\theta = \underline{\theta}$ and $\mu_w < \tilde{\mu}$, we have $p(x_1, \underline{\theta}, \emptyset) = \overline{p}$. Finally, $p(x_1, \underline{\theta}, \theta) = \overline{p}$. Combining these terms yields the result.

(ii) $x^M > \overline{x}$. By Lemma 2, $a_1^* = 0$. By Lemma 1, $\mu \leq \tilde{\mu}$ whenever $\theta = \underline{\theta}$. By (6), $s^* = M$ if and only if $\mu \leq \tilde{\mu}$. Thus $g^M(\underline{\theta}, e) = (1 + \delta)m$. If $\theta = \overline{\theta}$, then M receives $(1 + \delta)m$ if $\mu \leq \tilde{\mu}$, and m otherwise. Hence $\min_e g^M(\overline{\theta}, e) \leq \max_e g^M(\underline{\theta}, e)$, so M weakly prefers lower effort.

First I establish the optimal punishment strategy that conditions on θ and x_1. Since $x_1 = q$, $p(\cdot)$ can only condition on θ. Any punishment strategy is thus characterized by a pair $(p(x_1, \overline{\theta}, w), p(x_1, \underline{\theta}, w)) \in [0, \overline{p}]^2$. E's objective is then:

$$\begin{aligned} U^E = {} & eg^E(\overline{\theta}, e^*) + (1-e)g^E(\underline{\theta}, e^*) - ep(x_1, \overline{\theta}, w) \\ & - (1-e)p(x_1, \underline{\theta}, w) - ce^2. \end{aligned}$$

This objective is concave in e. Clearly, effort is minimized if $\frac{dU^E}{de}$ is minimized, and $\frac{dU^E}{de}$ is minimized if $p(x_1, \overline{\theta}, w) - p(x_1, \underline{\theta}, w)$ is maximized. Thus the optimal punishment strategy is $p(x_1, \overline{\theta}, w) = \overline{p}$ and $p(x_1, \underline{\theta}, w) = 0$ for all x_1 and w.

The optimal punishment strategy that also conditions on w is derived in a manner analogous to that in part (i) and is therefore omitted. ■

Proof of Proposition 1. First I derive the effort levels induced under each punishment scheme, taking $\tilde{x}$ as given. There are three cases. First, if $c_w > \delta l(\underline{\theta})$, then whistleblowing is dominated and M punishes based on quality (14). E's objective is then:

$$\begin{aligned} U^E = {} & e(1 + \delta)\left[b^E\int_X x f(x \mid \overline{\theta})\,dx - k^E\right] \\ & + (1-e)\left[1 + \delta(1 - F(\tilde{x} \mid \underline{\theta}))\right]\left[b^E\int_X x f(x \mid \underline{\theta})\,dx - k^E\right] \\ & - (1-e)\overline{p} - ce^2 \end{aligned}$$

https://doi.org/10.1017/S0003055408080192 Published online by Cambridge University Press

$$= (1+\delta)[b^E(e\overline{x}+(1-e)\underline{x})-k^E] - \delta(1-e)F(\tilde{x}\,|\,\underline{\theta})(b^E\underline{x}-k^E)-(1-e)\overline{p}-ce^2.$$

This objective is concave. Straightforward optimization yields:

$$e_n^* = \frac{(1+\delta)b^E(\overline{x}-\underline{x})+\delta F(\tilde{x}\,|\,\underline{\theta})(b^E\underline{x}-k^E)+\overline{p}}{2c}. \quad \textbf{(18)}$$

Second, $c_w \le \delta l(\underline{\theta})$ but M punishes based on quality. Now Lemma 1 implies that $\mu > \tilde{\mu}\ (<\tilde{\mu})$ if $\theta = \overline{\theta}\ (=\underline{\theta})$. Thus, $a_2^* = 1$ iff $\theta = \overline{\theta}$. E's objective can now be written as:

$$\begin{aligned} U^E &= e(1+\delta)\left[b^E\int_X x f(x\,|\,\overline{\theta})\,dx - k^E\right] \\ &\quad +(1-e)\left[\left(b^E\int_X x f(x\,|\,\underline{\theta})\,dx\right)-k^E-(1-F(\tilde{x}\,|\,\underline{\theta}))c_w\right] \\ &\quad -(1-e)\overline{p}-ce^2. \\ &= b^E(e\overline{x}+(1-e)\underline{x})-k^E+\delta e(b^E\overline{x}-k^E) \\ &\quad -(1-e)(1-F(\tilde{x}\,|\,\underline{\theta}))c_w-(1-e)\overline{p}-ce^2. \end{aligned} \quad \textbf{(19)}$$

This objective is also concave, and straightforward optimization yields:

$$e_\theta^* = \frac{b^E(\overline{x}-\underline{x})+\delta(b^E\overline{x}-k^E)+(1-F(\tilde{x}\,|\,\underline{\theta}))c_w+\overline{p}}{2c}. \quad \textbf{(20)}$$

Third, $c_w \le \delta l(\underline{\theta})$ and M punishes based on whistleblowing (15). As in the first case, r^* (11) ensures that P always correctly infers that $\theta = \overline{\theta}$, but not $\theta = \underline{\theta}$. E's objective is thus:

$$\begin{aligned} U^E &= e(1+\delta)\left[b^E\int_X x f(x\,|\,\overline{\theta})\,dx - k^E\right] \\ &\quad +(1-e)\left[1+\delta(1-F(\tilde{x}\,|\,\underline{\theta}))\right]\left[b^E\int_X x f(x\,|\,\underline{\theta})\,dx - k^E\right] \\ &\quad -(1-e)\left[(1-F(\tilde{x}\,|\,\underline{\theta}))(\overline{p}-\delta l(\underline{\theta})+c_w)+F(\tilde{x}\,|\,\underline{\theta})\overline{p}\right]-ce^2 \\ &= (1+\delta)[b^E(e\overline{x}+(1-e)\underline{x})-k^E]-\delta(1-e)(b^E\underline{x}-k^E) \\ &\quad -(1-e)(1-F(\tilde{x}\,|\,\underline{\theta}))c_w-(1-e)\overline{p}-ce^2. \end{aligned} \quad \textbf{(21)}$$

Comparing (19) with (21) reveals that the two objectives are identical, and thus the effort level for punishing based on whistleblowing is also e_θ^*.

To show which punishment strategy and effort levels are adopted in equilibrium, note first that if $c_w > \delta l(\underline{\theta})$, then M clearly uses $p_\theta^*(\cdot)$ (14) and $e^* = e_n^*$. Otherwise, if $\overline{p} < \delta l(\underline{\theta}) - c_w$, then $p_w^*(\cdot)$ (15) is infeasible and M uses $p_\theta^*(\cdot)$ and $e^* = e_\theta^*$. Finally, if $\overline{p} \ge \delta l(\underline{\theta}) - c_w$, then since effort and a_1^* are identical under both strategies, it is sufficient to compare M's period 2 payoffs under each. The punishment strategy $p_w^*(\cdot)$ is then superior to $p_\theta^*(\cdot)$ if:

$$\begin{aligned} &e_\theta^*(b^M\overline{x}-k^M+m)+(1-e_\theta^*)[(1-F(\tilde{x}))(b^M\underline{x}-k^M+m)] \\ &\quad \ge e_\theta^*(b^M\overline{x}-k^M+m) \\ &\quad \Leftrightarrow (1-e_\theta^*)(1-F(\tilde{x}))(b^M\underline{x}-k^M+m)\ge 0. \end{aligned}$$

Since M is aggressive, the left-hand side of the last expression is always non-negative, and hence M uses $p_w^*(\cdot)$ and $e^* = e_\theta^*$.

Finally I show the existence of a standard $\tilde{x}$ at which P's posterior belief is $\mu_r < (>)\ \tilde{\mu}$ for $x_1 < (>)\ \tilde{x}$. Let $e^*(x)$ be the optimal effort implied by a cutoff at any $x \in X$, and let $\mu_r(x)$ be the associated posterior belief. Applying (4) and (13), $\mu_r(x) = \tilde{\mu}$ is equivalently:

$$\frac{f(x|\overline{\theta})e^*(x)}{f(x|\overline{\theta})e^*(x)+f(x|\underline{\theta})(1-e^*(x))} = \frac{k^P/b^P-\underline{x}}{\overline{x}-\underline{x}}. \quad \textbf{(22)}$$

By (3), $\tilde{\mu} \in (0,1)$. By the continuity of $f(\cdot)$, $e^*(\cdot)$ is continuous. Therefore $\mu_r(x)$ is continuous in x and $\mu_r(x) \in [0,1]$. Now $\tilde{x}$ can be uniquely defined as follows:

$$\tilde{x} = \begin{cases} \min X & \text{if } \mu_r(x) > \tilde{\mu} \text{ for all } x \\ \max X & \text{if } \mu_r(x) < \tilde{\mu} \text{ for all } x \\ \min\{x \mid (22) \text{ holds}\} & \text{otherwise.} \end{cases}$$

To show that $\tilde{x}$ has the desired properties, observe that by MLRP (2), μ_r is increasing in x_1. Thus given effort $e^*(\tilde{x})$, we have $\mu_r < (>)\ \tilde{\mu}$ for $x_1 < (>)\ \tilde{x}$. ■

Proof of Proposition 2. First I derive the effort levels induced under each punishment scheme. Note throughout that $a_1^* = 0$ and Bayes' Rule implies $\mu_r = e^*$. There are three cases. First, if $c_w > \delta l(\overline{\theta})$, then whistleblowing is dominated and M punishes based on quality (14). There are two possible solutions. If $e^* > \tilde{\mu}$, then r^* (11) implies $\mu = 1$, $s^* = P$ and $a_2^* = 1$ if $\theta = \overline{\theta}$; and $\mu = 0$, $s^* = M$ and $a_2^* = 0$ if $\theta = \underline{\theta}$. E's objective is then:

$$U^E = \delta e(b^E\overline{x}-k^E)-e\overline{p}-ce^2. \quad \textbf{(23)}$$

This objective is concave, so differentiating and solving produces:

$$e_n = \frac{\delta(b^E\overline{x}-k^E)-\overline{p}}{2c}. \quad \textbf{(24)}$$

This is an equilibrium effort level if $e_n > \tilde{\mu}$.

Likewise, if $e^* \le \tilde{\mu}$, then $s^* = M$, $a_2^* = 0$ for all θ, and E's objective is $U^E = -e\overline{p}-ce^2$, which is also concave. Since $e \ge 0$, this yields a corner solution at 0, which is less than $\tilde{\mu}$. Although this corner solution always exists, the equilibrium that maximizes E's effort is selected, and so the equilibrium effort level is:

$$e_n^* = \begin{cases} e_n & \text{if } e_n > \tilde{\mu} \\ 0 & \text{otherwise.} \end{cases} \quad \textbf{(25)}$$

Second, $c_w \le \delta l(\overline{\theta})$ but M punishes based on quality. Again there are two possible solutions. If $e^* > \tilde{\mu}$, then the derivation is identical to the previous case and the solution is e_n. If $e^* \le \tilde{\mu}$, then Lemma 1 implies that $w^* = \theta$ iff $\theta = \overline{\theta}$ and $r^* = \emptyset$. Thus, r^* (11) implies $a_2^* = 1$ iff $\theta = \overline{\theta}$. E's objective can now be written as:

$$U^E = \delta e(b^E\overline{x}-k^E)-e(\overline{p}+c_w)-ce^2. \quad \textbf{(26)}$$

This objective is concave, and straightforward optimization yields the optimum effort level of $\frac{\delta(b^E\overline{x}-k^E)-c_w-\overline{p}}{2c}$ or 0 at a corner solution (where $\overline{p} \ge \delta l(\overline{\theta}) - c_w$). Combining cases we have:

$$e_\theta^* = \begin{cases} e_n & \text{if } e_n > \tilde{\mu} \\ \max\left\{0, \frac{\delta(b^E\overline{x}-k^E)-c_w-\overline{p}}{2c}\right\} & \text{otherwise.} \end{cases} \quad \textbf{(27)}$$

Third, $c_w \le \delta l(\overline{\theta})$ and M punishes based on whistleblowing (15), which requires $\overline{p} \ge \delta l(\overline{\theta}) - c_w$. Again there are two subcases. In the first, $e^* > \tilde{\mu}$. Since $r^* = \theta$ if $\theta = \underline{\theta}$, Bayes' Rule implies $\mu_w < \tilde{\mu}\ (>\tilde{\mu})$ if $\theta = \underline{\theta}\ (=\overline{\theta})$. Thus, $w^* = \emptyset$. With no whistleblowing to deter, M must punish based on quality. E's objective is then identical to (23) and the optimal effort is e_n. In the second, $e^* \le \tilde{\mu}$ because $e_n \le \tilde{\mu}$. Now $s^* = M$ unless

https://doi.org/10.1017/S0003055408080192 Published online by Cambridge University Press

$\theta = \overline{\theta}$ and $w = \theta$. But since whistleblowing is deterred, E's objective becomes $U^E = -e(\overline{p} - \delta l(\overline{\theta}) + c_w) - ce^2$. Solving for this subcase yields the corner effort level of 0. Thus we have the following effort levels:

$$e_w^* = e_n^* = \begin{cases} e_n & \text{if } e_n > \tilde{\mu} \\ 0 & \text{otherwise.} \end{cases} \tag{28}$$

To see which punishment strategy M adopts in equilibrium, observe first that if $c_w > \delta l(\overline{\theta})$, then M clearly uses $p_\theta^*(\cdot)$ (14) and $e^* = e_n^*$. Now let $c_w \leq \delta l(\overline{\theta})$. If $\overline{p} < \delta l(\overline{\theta}) - c_w$, then nonmaximal punishments for $\overline{\theta}$ are dominated and M again uses $p_\theta^*(\cdot)$; thus, $e^* = e_\theta^*$.

If $\overline{p} \geq \delta l(\overline{\theta}) - c_w$, there are two cases. First, if $e_n > \tilde{\mu}$, then $e_w^* = e_n^* = e_\theta^*$. If M punishes based on whistleblowing, then since $x_1 = q$ and $\mu_r = e_w^*$, in equilibrium $\mu > \tilde{\mu}$ unless M reports θ. By Lemma 1, this occurs when $\theta = \underline{\theta}$, and so $\mu > \tilde{\mu}$ $(< \tilde{\mu})$ when $\theta = \overline{\theta}$ $(= \underline{\theta})$. Likewise, by Lemma 1, when M punishes based on quality, $\mu > \tilde{\mu}$ $(< \tilde{\mu})$ when $\theta = \overline{\theta}$ $(= \underline{\theta})$. Thus under both punishment strategies, effort is identical, $w = \emptyset, \theta$ is revealed and P's responses are identical. Therefore $e^* = e_n$ and $p^*(\cdot) = p_\theta^*(\cdot)$. Second, if $e_n \leq \tilde{\mu}$, then $e_w^* = e_\theta^* = 0$. Thus under either punishment strategy, $\theta = \underline{\theta}$ with certainty and M receives a second period payoff of m. Therefore $e^* = 0$ and $p^*(\cdot) = p_\theta^*(\cdot)$.

Combining the expressions yields the result. ■

Proof of Comment 1. (i) If M is aggressive, then by Proposition 1 P's expected utility is:

$$U^P = \begin{cases} e_\theta^*(1+\delta)(b^P\overline{x} - k^P) + (1-e_\theta^*)(b^P\underline{x} - k^P) & \text{if } c_w < \delta l(\overline{\theta}) - \overline{p} \quad (a) \\ e_\theta^*(1+\delta)(b^P\overline{x} - k^P) + (1-e_\theta^*)[1+\delta(1-F(\tilde{x}\,|\,\underline{\theta}))](b^P\underline{x} - k^P) & \text{if } c_w \in [\delta l(\overline{\theta}) - \overline{p}, \delta l(\overline{\theta})] \quad (b) \\ e_n^*(1+\delta)(b^P\overline{x} - k^P) + (1-e_n^*)[1+\delta(1-F(\tilde{x}\,|\,\underline{\theta}))](b^P\underline{x} - k^P) & \text{if } c_w > \delta l(\overline{\theta}) \quad (c) \end{cases} \tag{29}$$

where e_θ^* is given by (20) and e_n^* is given by (18). To derive a condition under which $c_w > \delta l(\overline{\theta})$ is optimal for P, it is sufficient to compare the expected utility in case (c) with those of cases (a)–(b).

It is easily verified from Proposition 1 that $e_\theta^* \leq e_n^*$. Then manipulating cases (a) and (c), U^P is higher when $c_w > \delta l(\overline{\theta})$ than when $c_w < \delta l(\overline{\theta}) - \overline{p}$ if:

$$\delta \leq \delta^* \equiv \begin{cases} -\frac{(e_n^* - e_\theta^*)b^P(\overline{x} - \underline{x})}{(e_n^* - e_\theta^*)(b^P\overline{x} - k^P) + (1 - F(\tilde{x}\,|\,\underline{\theta}))(1 - e_n^*)(b^P\underline{x} - k^P)} & \text{if } (e_n^* - e_\theta^*)(b^P\overline{x} - k^P) + (1 - F(\tilde{x}\,|\,\underline{\theta}))(1 - e_n^*)(b^P\underline{x} - k^P) < 0 \\ \infty & \text{otherwise.} \end{cases}$$

Note that by assumption (3), $b^P\overline{x} - k^P > 0 > b^P\underline{x} - k^P$. Combined with the fact that $e_\theta^* \leq e_n^*$, it is clear that δ^* is nonnegative.

The comparison of expected utility for $c_w > \delta l(\overline{\theta})$ and $c_w \in [\delta l(\overline{\theta}) - \overline{p}, \delta l(\overline{\theta})]$ cases (b) and (c) of (29) is almost identical and is therefore omitted.

(ii) It is straightforward to verify that e_θ^* is continuous in c_w over a neighborhood of $c_w = \delta l(\overline{\theta}) - \overline{p}$. It is then sufficient to compare U^P in cases (a) and (b) of (29) at $c_w = \delta l(\overline{\theta}) - \overline{p}$. Simplifying yields the condition that U^P is higher in case (a) if $0 > (1 - e^*)\delta(1 - F(\tilde{x}\,|\,\underline{\theta}))(b^P\underline{x} - k^P)$. The requirement that $\tilde{x} \in \text{int}\, X$ implies $e^* < 1$ and $F(\tilde{x}|\underline{\theta}) < 1$. Additionally, by assumption (3), $b^P\underline{x} - k^P < 0$. Thus the condition holds.

(iii) If M is conservative, then by Proposition 2 P's expected utility is:

$$U^P = \begin{cases} \delta\frac{\delta(b^E\overline{x} - k^E) - \overline{p}}{2c}(b^P\overline{x} - k^P) & \text{if } \frac{\delta(b^E\overline{x} - k^E) - \overline{p}}{2c} \geq \tilde{\mu} \quad (a) \\ \delta\frac{\delta(b^E\overline{x} - k^E) - c_w - \overline{p}}{2c}(b^P\overline{x} - k^P) & \text{if } \frac{\delta(b^E\overline{x} - k^E) - \overline{p}}{2c} < \tilde{\mu} \text{ and } c_w < \delta l(\overline{\theta}) - \overline{p} \quad (b) \\ 0 & \text{if } \frac{\delta(b^E\overline{x} - k^E) - \overline{p}}{2c} < \tilde{\mu} \text{ and } c_w \geq \delta l(\overline{\theta}) - \overline{p} \quad (c). \end{cases} \tag{30}$$

It is thus clear that U^P is constant in c_w if $\frac{\delta(b^E\overline{x} - k^E) - \overline{p}}{2c} \geq \tilde{\mu}$ (case (a)), and weakly decreasing in c_w otherwise (cases (b)–(c)). ■

Proof of Comment 2. Since $\overline{p}^w$ limits punishments when $w = \theta$, it is clear that the result holds for any $p^*(x_1, \theta, \theta)$. Now suppose that $p^*(x_1', \theta', \emptyset) > \overline{p}^w + c_w$ for some x_1' and θ'. Upon the realization of x_1' and θ', E can choose $w = \theta$ and the punishment will be some $p^*(x_1', \theta', \theta) = p' \leq \overline{p}^w$. By revealing θ, E may not benefit only if $x^E < \underline{x}$ and $\theta = \underline{\theta}$, or $x^E > \overline{x}$ and $\theta = \overline{\theta}$. In the former case, by assumption $x^M > \underline{x}$ (M is conservative). It is then clear from Lemma 1 that M's report r ensures that $\mu < \tilde{\mu}$ when $\theta = \underline{\theta}$, and thus w cannot change P's response s. A symmetrical argument holds for the latter (aggressive M) case. Thus E chooses $w = \theta$ when $\theta = \theta'$ and $x_1 = x_1'$.

M can therefore replace this punishment with $p(x_1', \theta', \emptyset) = p' + c_w$. Now if $w = \theta$, M receives the same payoff as under $p^*(\cdot)$. If $w = \emptyset$, M receives at least as high of a payoff as under $p^*(\cdot)$, since M could have received the payoff from $p^*(\cdot)$ by reporting $r = \theta$. Thus M cannot benefit from any punishment strategy where $p^*(x_1', \theta', \emptyset) > \overline{p}^w + c_w$. ■

Proof of Comment 3. (i) If M is aggressive and $\overline{p} < \delta l(\underline{\theta}) - c_w$, then P's expected utility is given by case (a) of (29). Differentiating with respect to e_θ^* yields $\frac{dU^P}{de_\theta^*} = (1+\delta)(b^P\overline{x} - k^P) - (b^P\underline{x} - k^P)$. This is positive by assumption (3), and thus it is sufficient to show that e_θ^* is increasing in $\overline{p}$.

Let $u'(e; \overline{p}) = b^E(\overline{x} - \underline{x}) + \delta(b^E\overline{x} - k^E) + [(1 - e)\frac{dF(\tilde{x}\,|\,\underline{\theta})}{d\tilde{x}}\frac{d\tilde{x}}{de} + (1 - F(\tilde{x}\,|\,\underline{\theta}))]c_w + \overline{p} - 2ce$ denote the first order condition on E's utility (adapted from (19)). Applying the Implicit Function Theorem, we have $\text{sign}(\frac{de^*}{d\overline{p}}) = \text{sign}(\frac{\partial u'}{\partial \overline{p}}(e^*(\overline{p}); \overline{p}))$. Since the latter derivative evaluates to 1, U^P must be increasing in $\overline{p}$ for $\overline{p} < \delta l(\underline{\theta}) - c_w$.

(ii) This result is equivalent to Comment 1(ii).

(iii) If M is conservative, then P's expected utility is given by (29). The result follows from three observations. First, all values of U^P in case (a) exceed all values of U^P in case (b), which in turn exceed all values of U^P in case (c). Second, U^P is strictly decreasing in $\overline{p}$ in cases (a)–(b), and constant in case (c). Finally, cases (a), (b) and (c) partition values of $\overline{p}$ from lowest to highest. ■

https://doi.org/10.1017/S0003055408080192 Published online by Cambridge University Press

REFERENCES

Alford, C. Fred. 2001. *Whistleblowers: Broken Lives and Organizational Power.* Princeton: Princeton University Press.

Apesteguia, Jose, Martin Dufwenberg, and Reinhard Selten. 2007. "Blowing the Whistle." *Economic Theory* 31: 143–66.

Aubert, Cécile, Patrick Rey, and William E. Kovacic. 2006. "The Impact of Leniency and Whistle-blowing Programs on Cartels." *International Journal of Industrial Organization* 24: 1241–66.

Banks, Jeffrey S. 1989. "Agency Budgets, Cost Information, and Auditing." *American Journal of Political Science* 33 (3): 670–99.

Bendor, Jonathan B. 1985. *Parallel Systems: Redundancy in Government.* Berkeley, CA: University of California Press.

Bolton, Patrick, and Mathias Dewatripont. 1994. "The Firm as a Communication Network." *Quarterly Journal of Economics* 109 (4): 809–39.

Bowman, James S. 1983. "Whistle Blowing: Literature and Resource Materials." *Public Administration Review* 43 (3): 271–76.

Carpenter, Daniel P. 2004. "The Political Economy of FDA Drug Approval: Processing, Politics and Lessons for Policy." *Health Affairs* 23 (1): 52–63.

Carpenter, Daniel P., and Michael M. Ting. 2007. "Regulatory Errors with Endogenous Agendas." *American Journal of Political Science* 51 (4): 835–52.

De Maria, William. 1999. *Whistleblowing and the Ethical Meltdown of Australia.* Kent Town, Australia: Wakefield Press.

Dixit, Avinash K. 1998. *The Making of Economic Policy: A Transaction Cost Politics Perspective.* Cambridge: MIT Press.

Dixit, Avinash K. 2002. "Incentives and Organizations in the Public Sector: An Interpretative Review." *Journal of Human Resources* 37 (4): 696–727.

Epstein, David, and Sharyn O'Halloran. 1994. "Administrative Procedures, Information, and Agency Discretion." *American Journal of Political Science* 39 (3): 697–722.

Fayol, Henri. 1949. *General and Industrial Management.* London: Pitman.

Friebel, Guido, and Michael Raith. 2004. "Abuse of Authority and Hierarchical Communication." *Rand Journal of Economics* 35 (2): 224–44.

Gailmard, Sean. 2007. "Multiple Principals and Oversight of Bureaucratic Policy-Making." Unpublished manuscript, University of California at Berkeley.

Gailmard, Sean, and John W. Patty. 2007. "Slackers and Zealots: Civil Service, Policy Discretion and Bureaucratic Expertise." *American Journal of Political Science* 51 (4): 873–89.

Gibbons, Robert. 1998. "Incentives in Organizations." *Journal of Economic Perspectives* 12 (4): 115–32.

Heclo, Hugh. 1977. *A Government of Strangers: Executive Politics in Washington.* Washington, DC: The Brookings Institution.

Heimann, C. F. Larry. 1993. "Understanding the Challenger Disaster: Organizational Structure and the Design of Reliable Systems." *American Political Science Review* 87 (2): 421–35.

Heimann, C. F. Larry. 1997. *Acceptable Risks: Politics, Policy, and Risky Technologies.* Ann Arbor: The University of Michigan Press.

Holmstrom, Bengt, and Paul Milgrom. 1991. "Multitask Principal Agent Analyses: Incentive Contracts, Asset Ownership and Job Design." *Journal of Law, Economics, and Organization* 7 (Special Issue): 24–52.

Hopenhayn, Hugo, and Susanne Lohmann. 1996. "Fire-Alarm Signals and the Political Oversight of Regulatory Agencies." *Journal of Law, Economics, and Organization* 12 (1): 196–213.

Knott, Jack H., and Gary J. Miller. 1987. *Reforming Bureaucracy: The Politics of Institutional Choice.* Englewood Cliffs, NJ: Prentice-Hall.

Kovacic, William E. 1998. "The Civil False Claims Act as a Deterrent to Participation in Government Procurement Markets." *Supreme Court Economic Review* 6: 201–39.

Krause, George A., David E. Lewis, and James W. Douglas. 2006. "Political Appointments, Civil Service Systems, and Bureaucratic Competence: Organizational Balancing and Executive Branch Revenue Forecasts in the American States." *American Journal of Political Science* 50 (3): 770–87.

Laffont, Jean-Jacques. 1990. "Analysis of Hidden Gaming in a Three-Level Hierarchy." *Journal of Law, Economics, and Organization* 6 (2): 301–24.

Lewis, David. 2001. "Whistleblowing at Work: On What Principles Should Legislation Be Based?" *Industrial Law Journal* 30 (2): 169–93.

Lewis, David E. 2003. *Presidents and the Politics of Agency Design.* Stanford, CA: Stanford University Press.

Lewis, David E. 2008. *The Politics of Presidential Appointments: Political Control and Bureaucratic Performance.* Princeton: Princeton University Press.

Lorentzen, Peter L. 2007. "Regularized Rioting: The Strategic Toleration of Popular Protest in China." Unpublished manuscript, University of California at Berkeley.

McAfee, R. Preston, and John McMillan. 1995. "Organizational Diseconomies of Scale." *Journal of Economics and Management Strategy* 4: 399–426.

McCubbins, Mathew D., Roger G. Noll, and Barry R. Weingast. 1987. "Administrative Procedures as Instruments of Political Control." *Journal of Law, Economics, and Organization* 3 (2): 243–77.

McCubbins, Mathew D., and Thomas Schwartz. 1984. "Congressional Oversight Overlooked: Police Patrols Versus Fire Alarms." *American Journal of Political Science* 28 (1): 165–79.

Melumad, Nahum D., Dilip Mookherjee, and Stefan Reichelstein. 1995. "Hierarchical Decentralization of Incentive Contracts." *Rand Journal of Economics* 26 (4): 654–72.

Moe, Terry M. 1989. "The Politics of Bureaucratic Structure." In ed. John E. Chubb and Paul E. Peterson, *Can the Government Govern?* Washington, DC: The Brookings Institution.

Near, Janet, and Marcia Miceli. 1996. "Whistle-Blowing: Myth and Reality." *Journal of Management* 22 (3): 507–26.

Park, Valerie R. 1991. "The False Claims Act, Qui Tam Relators, and the Government: Which Is the Real Party to the Action?" *Stanford Law Review* 43 (5): 1061–93.

Prendergast, Canice. 2003. "The Limits of Bureaucratic Efficiency." *Journal of Political Economy* 111 (5): 929–58.

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

Shafritz, Jay M., and E. W. Russell. 2000. *Introducing Public Administration.* 2nd ed. New York: Addison Wesley Longman.

Ting, Michael M. 2002. "A Theory of Jurisdictional Assignments in Bureaucracies." *American Journal of Political Science* 46 (2): 364–78.

Tirole, Jean. 1986. "Hierarchies and Bureaucracies: On the Role of Collusion in Organizations." *Journal of Law, Economics, and Organization* 2 (2): 181–214.

United States Office of Special Counsel. 1999. "A Report to Congress from the U.S. Office of Special Counsel for Fiscal Year 1999."

United States Congressional Research Service. 2005. "National Security Whistleblowers." CRS Report RL33215.

United States Congressional Research Service. 2007. "The Whistleblower Protection Act: An Overview." CRS Report RL33918.

Wilson, James Q. 2000. *Bureaucracy: What Government Agencies Do and Why They Do It.* 2nd ed. New York: Basic Books.

https://doi.org/10.1017/S0003055408080192 Published online by Cambridge University Press