WHISTLEBLOWING*

Michael M. Ting†
Department of Political Science and SIPA
Columbia University

August 16, 2006

Abstract

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 discipline employees. Politicians must therefore strike a balance between information revelation and the preservation of bureaucrats' incentives to exert effort. This paper explores these tradeoffs with a model of agency decision-making under incomplete information. In the game, an employee's effort determines a project's type, and a manager chooses whether to approve the project and discipline the employee. By whistleblowing, an employee reveals the type to a politician, who may override the manager's decision. While whistleblowing always increases the transmission of information, its effects on employee effort depend on managerial preferences. A key finding is that stronger whistleblower protections reduce effort when the manager is "aggressive" and would commit more Type I errors than the politician would, but increase effort otherwise. Whistleblower protections therefore unambiguously benefit politicians if an agency is inclined to make Type II errors.

*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, and the University of Chicago. I thank Stuart Jordan, Patrick Warren, and Alan Wiseman for helpful comments, and Arne Grafweg, Caroline McGregor, and Andrea Venneri for research assistance.

†Political Science Department, 420 W 118th St., New York NY 10027 (mmt2033@columbia.edu).

1. Introduction

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 anti-terrorism 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 post-approval 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 U.S., 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 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] All of these laws have been amended (and usually strengthened) numerous times. Similar 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]

[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. The protections are generally weaker for employees in national security organizations; see Congressional Research Service report RL33215 (2005) for an overview. 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.

To date, there has been relatively little scrutiny of the effects of whistleblowing and whistleblowing policy on organizational performance.[4] However, given its importance for policing government agencies and private contractors, there exists something of a consensus today that whistleblowing and its legal protection are essential.[5] 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 (such as Congress) can therefore benefit from the information possessed by organization members (such as research staff) who do not normally interact with these principals.

While politicians can certainly benefit from revelations by whistleblowers, the argument for their protection runs directly counter to one of the classic intuitions of organization theory. 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 organizational performance by removing bad managerial incentives. A manager worried about being publicly exposed by a subordinate might divert effort toward suppressing employees, or might not select the best employees (*e.g.*, Friebel and Raith 2004).

A politician or voter interested in agency performance may therefore find *ex post* incentives for revealing information to be in tension with *ex ante* incentives for inducing effort. In particular, *ex post* incentives may dilute *ex ante* incentives because the choices available to a decision-maker are not independent of her subordinate's or agent's effort. For example, upon hearing a whistleblower a politician could conclude that a certain project should be terminated due to low quality. But that quality could be determined by employee actions that are uncontractable. In such cases the politician may wish not to restrict a manager's latitude to discipline employees. This tension is especially relevant in the public sector, where civil service protections heavily constrain the incentives that managers can provide for employee effort (*e.g.*, Knott and Miller 1987). In fact, several significant court cases have invoked this rough intuition. 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.[6]

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

[5]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).

[6]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

This paper develops a model that illuminates these basic incentives and generates implications for public sector whistleblowing policy. One of its key innovations is that it considers simultaneously an agency's internal structure as well as its external environment. Naturally, the model also has a number of limitations. First, it does not feature a policy dimension, and instead focuses on the implementation (or non-implementation) of a fixed project pursuant to some established law. The structure of the project corresponds with commonly observed bureaucratic decisions, for example approving a pharmaceutical product or launching a rocket. Second, it ignores transaction costs. One argument against elaborate whistleblower protections is that such procedures are costly and cumbersome. While these costs are certainly nontrivial in practice, the focus here will be restricted to the interaction between effort and information. Finally, it is concerned generally with organizational performance, and not with the protection of whistleblower rights *per se*.[7]

The model is a game with three players; a manager and an employee who form an organization, as well as a politician who monitors its behavior. This environment best describes a public agency, where the employee is a civil servant whose terms of employment cannot be altered, and the manager is a political appointee who can be replaced easily by the politician. It may also apply to voters and elected officials, or to shareholders and management in public corporations. The players contribute to the output of a single project that generates a publicly observable outcome in each of two periods. All players are interested in these outcomes, but have different levels of knowledge and ability to affect outcomes. Members of the organization are also motivated in part by the possibility of punishments from one level up. The employee may face punishment from the manager, while the manager may lose her decision-making authority to the politician.[8]

The game begins with the employee's choice of a costly and nonverifiable effort level. This effort probabilistically determines the project's "type," 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 the type is generated. Between periods, the manager and employee are each able to reveal the type to the politician. This report can be considered a form of expert testimony, which organization members can only provide voluntarily. In

efficient provision of public services." In a 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.

[7]Concerns about employee rights do animate much of the policy discussion about whistleblowing. 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]It is crucial for the model that the manager be in a position of providing incentives to employees. In an environment in which employees possess managerially relevant information but do not face such incentives, whistleblowing protections are trivially desirable to outside politicians (aside perhaps from the costs of such protections).

this context, whistleblowing is simply an employee report. Following the report(s), the manager can punish the employee. While 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 her to continue exercising approval authority. Intuitively, revocation places the manager's department in receivership. Thus whistleblowing renders the managerial decision contractable. Note that neither the politician nor the manager can manipulate the type or the employee's effort.

The results of the model reveal a relationship between *ex ante* and *ex post* incentives. The ability to whistleblow is always helpful to the politician *given* the project's type. However, whistleblowing also affects the project's type. A manager who might have used all of her punishment capacity on inducing effort might divert some of this capacity toward deterring whistleblowing. Further, an employee might face lower-powered incentives because the ability to whistleblow ameliorates the consequences of a bad project type. The desirability of these effects depends on the whether the manager is more prone to committing Type I or Type II errors, relative to the principal. If the manager is "aggressive," in the sense of wanting to approve more types than the principal would (*i.e.*, making Type I errors), then whistleblowing hurts employee effort. Somewhat surprisingly, if the manager is "conservative," in the sense of wanting to reject too many projects (*i.e.*, making Type II errors), then there is no tradeoff between *ex ante* and *ex post* incentives. Whistleblowing *raises* employee effort. It follows that whistleblowing policy should also depend on managerial preferences. A principal should desire stronger whistleblowing protections – reducing the scope of managerial punishments, or allowing employees to claim some of the manager's surplus – when managers are conservative. With an aggressive manager, the optimal policy may even be to disallow whistleblowing.

Due to its combination of moral hazard and signaling, this game draws upon two significant families of models of bureaucracies. The first considers the provision of incentives within organizations (Gibbons 1998, Dixit 2002, Gailmard and Patty 2004). Of particular relevance are models of multiple tasks (Holmstrom and Milgrom 1981, Ting 2002) and common agency (Dixit 1998, Wilson 2000, Gailmard 2003). 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 political appointees in the U.S. executive branch (Heclo 1977, Lewis 2003).

A second family of related theories 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 three-tier institutional structure in this paper is more closely approximated by work on administrative procedures and agency design (McCubbins, Noll, and Weingast 1987, Moe 1989).[9] While 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. They are relatively silent, however, on the role groups may play in influencing, as opposed to revealing, policy "type."

The paper proceeds as follows. The next section formally lays out the whistleblowing model. Section 3 begins by considering first a baseline case in which no whistleblowing is permitted, and then derives the main results of the full model. Section 4 considers the implications of whistleblowing policies, including limiting managerial retaliation against whistleblowing and allowing employees to claim part of the manager's surplus. Section 5 summarizes and concludes.

2. The Model

The game, labeled Γ^w, 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 *type* $\theta \in \{\underline{\theta}, \overline{\theta}\}$. Each x_t is drawn i.i.d. according to a probability density $f(x_t|\theta)$ satisfying

[9] A number of models of three-tier hierarchies examine the performance of alternative organizational forms, though primarily in an adverse selection context; see *e.g.* McAfee and McMillan (1995) and Melumad, Mookherjee, and Reichelstein (1995).

$f(x_t|\theta) > 0$ for all θ and $x_t \in \text{int}X$, and $f(x_t|\theta) = 0$ otherwise. The density functions also satisfy the following:

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

$$\lim_{x_t \downarrow \min X} \frac{f(x_t|\overline{\theta})}{f(x_t|\underline{\theta})} = 0. \tag{3}$$

Assumption (2) is the familiar Monotone Likelihood Ratio Property (MLRP), which ensures that higher observations of x_t are more likely to be associated with the high type. Assumption (3) is adopted simply to avoid a number of corner solutions.

The expected value of x_t given type $\overline{\theta}$ ($\underline{\theta}$) is $\overline{x}$ ($\underline{x}$), where the "high" type has a higher expected outcome: $\overline{x} > \underline{x}$. To avoid a number of uninteresting cases, I assume throughout that:

$$x^P \in (\underline{x}, \overline{x}). \tag{4}$$

Thus, P prefers the expected outcome of the high type to q, and prefers q to the low type.[10]

Players inside the agency also receive payoffs from non-policy sources. M values office-holding and receives a fixed benefit of $m > 0$ for each period in which she holds managerial control. Additionally, E can be "punished" by M, which results in a loss of $p \in [0, \overline{p}]$. This can correspond to a re-assignment, delayed promotion or perhaps the dismissal of a political appointee. Finally, E must exert a one-time effort which affects x_t. The effort level $e \in [0, 1]$ imposes a cost ce^2, where $c > \max\{0, [(1+\delta)b^E\overline{x} - \delta k^E + \overline{p}]/2\}$ 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 type, where $\Pr\{\theta = \overline{\theta}\} = e$. The type is initially observable to 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 $a_t = 1$, N 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 during this phase is as follows. All actions are observable unless otherwise noted.

- M issues a report $r \in \{\emptyset, \theta\}$.

[10] If $x^P > \overline{x}$, then P wishes to see both types of projects rejected. Likewise, if $x^P < \underline{x}$, then P wishes to see both types approved. In either case, P's need for managerial discretion is greatly reduced.

- E makes a *whistleblowing* decision $w \in \{\emptyset, \theta\}$.
- 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 or convey nothing 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 (presumably at some unmodeled cost), but cannot be compelled to do so.

The punishment imposes a cost p on E but is costless to M. This is the mechanism through which M provides performance incentives to the employee. Finally, the choice of s gives P the opportunity to "renegotiate" second period decision-making rights. If $s = P$, then P assumes the manager's role in choosing a_2. If $s = M$, then M retains control and the second period of project execution is identical to that of the first. Note that P may condition this decision on θ only if $r = \theta$ or $w = \theta$.

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_t and x_t) 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 $w : [0, 1] \times \{\underline{\theta}, \overline{\theta}\} \times H_1 \times \{0, \theta\} \to \{\emptyset, \theta\}$ strategies. M's strategy has mappings $a_1 : \{\underline{\theta}, \overline{\theta}\} \to \{0, 1\}$ and $a_2 : \{\underline{\theta}, \overline{\theta}\} \times H_2 \to \{0, 1\}$ specifying period 1 and period 2 approval decisions. It also has measurable mappings $r : \{\underline{\theta}, \overline{\theta}\} \times H_1 \to \{\emptyset, \theta\}$ and $p : \{\underline{\theta}, \overline{\theta}\} \times H_1 \times \{\emptyset, \theta\}^2 \to [0, \overline{p}]$ specifying her reporting and punishment decisions, respectively.

P's strategy $s : H_1 \times \{\emptyset, \theta\}^2 \to \{M, P\}$ identifies the assignment of period 2 managerial rights. Additionally, P has posterior beliefs $\mu : H_1 \times \{\emptyset, \theta\}^2 \to [0, 1]$ that $\theta = \overline{\theta}$, 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.

To reduce the number of equilibrium cases of the model, it is assumed that M and P break ties in favor of approving projects. Additionally, P breaks ties in favor of allowing M to retain managerial control ($s = M$). This assumption is consistent with a cost of intervening in management decisions.

The model has multiple equilibria, which fortunately do not generally alter the conclusions. However, to simplify the analysis, two equilibrium selection rules are adopted. The first addresses 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 non-trivial 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 $r, w = \emptyset$. 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 that a_1 was incorrect from P's perspective. By the symmetry of the revelation technology, it should be clear that part (ii) cannot affect information revelation or equilibrium payoffs. 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 Section 3.

3. Main Results

Two variants of the game are developed here. The first, labeled Γ^n, allows no whistleblowing and serves as a baseline for comparison. This variant thus corresponds to a world in which information transmission between employees and policy-makers is very difficult. This might occur if the civil service system does not have institutionalized mechanisms for handling whistleblowers, or (unmodeled) credibility problems render employee reports unverifiable. The second, labeled Γ^w, restores the employee's ability to blow the whistle. This section informally discusses player strategies, which are formally derived in the Appendix and used in Propositions 1-3, which characterize the equilibrium effort levels and punishments. To keep the notation for strategies manageable, I omit the strategies' dependencies on information sets throughout, except where necessary.

To begin, observe that given P's beliefs, the two games are identical starting from P's choice s of period 2 decision rights. Thus, these moves may be considered first, independently of whether the employee may whistleblow.

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:

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

For any history in which $s = P$ and neither E nor M reveal θ, $E[x_t|h_2] = \mu\overline{x} + (1-\mu)\underline{x}$. Otherwise, $E[x_t|h_2]$ will be $\underline{x}$ or $\overline{x}$. 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 the manager's incentive to induce performance by punishing the employee—will in turn affect the employee's effort level.

Principal's Assignment of Decision Rights. Immediately preceding period 2, P chooses s. Given (5), this choice depends on whether P wishes to override M's authority, which in turn depends on her posterior beliefs over whether $\overline{\theta}$ is sufficiently likely. Clearly, if either $r = \theta$ or $w = \theta$, then θ is known: $\mu = 1$ $(= 0)$ if $\theta = \overline{\theta}$ $(= \underline{\theta})$. Otherwise, μ is interior. In this case, P is indifferent between assuming control and letting M retain control if $b^P(\mu\overline{x} + (1-\mu)\underline{x}) - k^P = 0$, where $\tilde{\mu}$ is P's posterior belief that $\theta = \overline{\theta}$. This implies the following "cutoff" value of μ:

$$\tilde{\mu} = \frac{k^P/b^P - \underline{x}}{\overline{x} - \underline{x}}. \tag{6}$$

Note that by (4), $\tilde{\mu} \in (0,1)$. Values of μ below (above) $\tilde{\mu}$ imply that P prefers a policy of $a_2 = 0$ (1). Since P breaks ties in favor of retaining M control, her optimal assignment of decision rights is:

$$s^* = \begin{cases} M & \text{if } \mu \geq \tilde{\mu} \text{ and } x^M < \overline{x} \\ & \quad \text{or } \mu \leq \tilde{\mu} \text{ and } x^M > \underline{x} \\ P & \text{if } \mu < \tilde{\mu} \text{ and } x^M < \underline{x} \\ & \quad \text{or } \mu > \tilde{\mu} \text{ and } x^M > \overline{x}. \end{cases} \tag{7}$$

Expression (7) implies that when $x^M \in (\underline{x}, \overline{x})$, P always delegates decision-making authority to M. This is intuitive, as M's preferences are aligned with P's for all values of θ.

3.1. A Baseline Case: No Whistleblowing

In Γ^n, w is constrained to be $\emptyset$. While a bad project cannot be revealed through whistleblowing, M must be concerned with P inferring that $\theta = \underline{\theta}$ from a bad first-period outcome. M's punishment strategy therefore induces the maximum possible performance from E.

To develop the equilibrium, consider the remainder of the game sequence in reverse order. In order to emphasize differences between Γ^n and the subsequent whistleblowing game, parameters of interest are denoted with a superscript n.

Managerial Punishment. Since punishment is costless for M, any choice of p is optimal at the punishment stage. Clearly, however, M would like to use p to induce optimal effort from E. This punishment can condition only on the observables x_1 and θ, and so M must "allocate" $\overline{p}$ across their realizations. The optimal strategy is for M to focus exclusively on θ:

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

When $x^M < \overline{x}$, M punishes maximally for realizations of the low type because this generates the greatest incentive for E to choose a high effort level. Somewhat counter-intuitively, M punishes the *high* type when $x^M > \overline{x}$ because she is able to secure an outcome of q for the low type, but not

necessarily for the high type. In both cases, conditioning on x_1 does not work as well because both types can generate high values of x_1 with positive probability.

Managerial Report. Generally, M will report θ to reverse P's posterior beliefs and retain managerial control. By revealing θ, the report results in $\mu = 0$ (for $\underline{\theta}$) or $\mu = 1$ ($\overline{\theta}$). Recall that μ_r is P's posterior belief prior to the choice of r. There are three cases. First, when $x^M \in (\underline{x}, \overline{x})$, M does not need to report because P trusts her to choose the correct action. Second, when $x^M < \underline{x}$, M wishes to convince P that $\theta = \overline{\theta}$, which removes P's incentive to "fire" M in period 2. M has no incentive to report θ if $\mu_r \geq \tilde{\mu}$, as P infers that $\overline{\theta}$ is sufficiently likely and allows M to approve the project again in period 2. M then benefits regardless of the true type. Thus M will report θ only when $\theta = \overline{\theta}$ and $\mu_r < \tilde{\mu}$. Finally, when $x^M > \overline{x}$, M wishes to convince P that $\theta = \underline{\theta}$, as P would then allow M to cancel the project in period 2. M's strategy therefore mirrors the second case, with a report if $\mu_r > \tilde{\mu}$. The optimal reporting strategy is then:

$$r^* = \begin{cases} \theta & \text{if } \theta = \overline{\theta}, \mu_r < \tilde{\mu}, \text{and } x^M < \underline{x}, \text{ or} \\ & \quad \theta = \underline{\theta}, \mu_r > \tilde{\mu}, \text{and } x^M > \overline{x} \\ \emptyset & \text{otherwise.} \end{cases} \tag{9}$$

The effect of this reporting strategy is to make P informed about θ whenever M and P agree on the approval action to be taken, conditional upon θ. P is "deceived" about θ, however, if the realization of x_1 suggests agreement when in fact there is none.[11]

Period 1 Approval. M has myopic incentives to approve or reject the project in a manner analogous to (5). However, she may wish to do the reverse to transmit information about θ to P. Clearly, this would not occur when $x^M \in (\underline{x}, \overline{x})$, as M and P have identical state-dependent preferences and P would always have correct posterior beliefs under M's reporting strategy. Less obviously, when $x^M \notin (\underline{x}, \overline{x})$, P's ability to learn θ when she agrees with M on the correct action also eliminates strategic approval by M. For example, when $x^M < \underline{x}$, a "separating" strategy of approving only when $\theta = \overline{\theta}$ would ensure that $\mu = 1$ whenever $x_t \neq q$. P then would not revoke M's authority when $\theta = \overline{\theta}$. However, (9) implies that P learns of the high type regardless of the approval strategy, while under the separating strategy M loses the ability to benefit from lucky realizations of μ when $\theta = \underline{\theta}$. M therefore never deviates from her myopic strategy:

$$a_1^* = \begin{cases} 1 & \text{if } b^M E[x_1|\theta] - k^M \geq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

[11] It is worth noting the role played by the "minimum reporting" refinement here. There are equilibria in which M reports θ when $\theta = \overline{\theta}$ for any subset of $\{x_1 \mid x_1 > \tilde{x}^n\}$. Given that P expects truthful reporting for any such x_1, P would infer silence by M as the low type, 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 $\theta = \underline{\theta}$.

This result simplifies the equilibrium characterization greatly by simplifying P's inferences from a_1 (or, equivalently, an observation of q). If $x^M \in (\underline{x}, \overline{x})$, then a_1 is completely informative and P knows θ with certainty. Otherwise, when $x^M \notin (\underline{x}, \overline{x})$, a_1 is completely uninformative and P does not use it to calculate her posterior beliefs. Thus P's beliefs, prior to M's report r, are:

$$\mu_r = \begin{cases} 1 & \text{if } x^M \in (\underline{x}, \overline{x}) \text{ and } a_1 = 1 \\ 0 & \text{if } x^M \in (\underline{x}, \overline{x}) \text{ and } a_1 = 0 \\ \frac{f(x_1|\overline{\theta})e}{f(x_1|\overline{\theta})e + f(x_1|\underline{\theta})(1-e)} & \text{otherwise.} \end{cases} \tag{11}$$

Employee Effort. The employee balances the cost of effort and the probability distribution over outcomes induced by that effort. Whether θ is revealed by M's subsequent action is a key factor in E's strategy. When all project types are approved at $t = 1$, the MLRP property (2) implies that μ is increasing in x_1. Thus there is a cutoff standard for x_1, $\tilde{x}^n \in X$, below (above) which P infers that $\mu < (>)\ \tilde{\mu}$, which determines her subsequent assignment of managerial authority. Because μ is calculated using Bayes' Rule, this standard must be consistent with E's effort level, as well as the realization of x_1.

The following result establishes the existence of $\tilde{x}^n$, and characterizes effort levels induced by different kinds of managers.

Proposition 1 *Principal's Standard and Employee Effort without Whistleblowing. There exists some $\tilde{x}^n \in X$ such that $\mu < (>)\ \tilde{\mu}$ for $x_1 < (>)\ \tilde{x}$. E's effort is $e^{n*} = \max\{0, e^n\}$, where:*

$$e^n = \begin{cases} \frac{(1+\delta)b^E(\overline{x}-\underline{x}) + \delta F(\tilde{x}^n|\underline{\theta})(b^E\underline{x} - k^E) + \overline{p}}{2c} & \text{if } x^M < \underline{x} \quad (i) \\ \frac{(1+\delta)(b^E\overline{x} - k^E) + \overline{p}}{2c} & \text{if } x^M \in (\underline{x}, \overline{x}) \quad (ii) \\ \frac{\delta(b^E\overline{x} - k^E) - \overline{p}}{2c} & \text{if } x^M > \overline{x} \text{ and } \frac{\delta(b^E\overline{x} - k^E) - \overline{p}}{2c} \geq \tilde{\mu} \quad (iii) \\ 0 & \text{if } x^M > \overline{x} \text{ and } \frac{\delta(b^E\overline{x} - k^E) - \overline{p}}{2c} < \tilde{\mu}. \quad (iv) \end{cases}$$

Proof. All proofs are in the Appendix. ■

The principal's standard $\tilde{x}^n$ appears only in Proposition 1(i), where the manager is "aggressive" and wishes to approve all projects. In this case, E must weigh the effect of e on x_1 reaching $\tilde{x}^n$. Note that a higher standard induces *lower* effort, since E prefers receivership to continued control by M when $\theta = \underline{\theta}$, and therefore faces a smaller downside risk when the cutoff is high.

In the other cases, P does not use x_1 to infer θ, and all players' period 2 payoffs do not depend on x_1. In case (ii), M is a perfect agent of P and so θ can be inferred perfectly. Cases (iii)-(iv) describe a "conservative" manager who wishes to reject all projects, and does so in equilibrium.

Even though $x_1 = q$, P can infer θ imperfectly because her beliefs are then exactly $\mu = e^*$. If $\mu \geq \tilde{\mu}$ (case (iii)), then P assumes control and approves the project in period 2. This case requires that E be interested in a high level of output, and thus $x^E < \overline{x}$. By contrast, if $\mu < \tilde{\mu}$ (case (iv)), then P allows M to continue to reject the project. The effort level is a corner solution at zero because given M cancellation in period 2, E's payoff from any $e > 0$ would be strictly negative.[12]

The effect of M's punishment for a particular type realization is to shift effort by $\overline{p}/2c$, except in case (iv). The direction of the shift is determined by whether M wishes for high (cases (i)-(ii)) or low (case (iii)) effort levels.

The game without whistleblowing generates some intuitive predictions. The manager is free to discipline the employee in in order to induce optimal performance, and consequently the probability of a good project is maximized when the manager is aggressive, as in Proposition 1(i). As is standard in many signaling games, information about the project is transmitted effectively when the manager's and principal's state-dependent preferences coincide, and less so when they do not. Finally, managerial authority is sometimes revoked in equilibrium, when the project is bad *and* the period 1 outcome suggests that the manager would choose the wrong action in period 2. The following figure illustrates the consequences for information revelation and managerial retention in the aggressive manager case.

[Figure 1]

3.2. The Whistleblowing Game

The full game, Γ^w, restores E's ability to report θ even when M does not. As a result, P may gain an additional opportunity to revoke M's decision rights in period 2. It is easy to see that when x^M and x^E are in the same interval relative to $\underline{x}$ and $\overline{x}$, this ability is inconsequential, and the equilibrium of Γ^n is unchanged. In other cases, however, M may choose to dissuade whistleblowing by conditioning her punishment on it.

Again, the period 2 approval and decision rights are determined by (5) and (7), so the analysis here begins with M's punishment strategy. Analogously with the previous game, parameters of interest here are denoted with a superscript w.

Managerial Punishment. To see the intuition for M's punishment strategy, it will be convenient to define:

$$l(\theta) = \left| b^E \int_X x f(x|\theta) dx - k^E \right| \tag{12}$$

[12]The zero-effort equilibrium also exists under the conditions of case (iii), but the equilibrium selection rule from Section 2 picks the equilibrium rule with the highest effort level.

as the difference between E's expected payoff in a single period from a type θ policy and zero (*i.e.*, the payoff from outcome q). By whistleblowing, E essentially hopes to gain $\delta l(\theta)$ from a change in managerial authority. M can therefore *prevent* whistleblowing — essentially, issuing a "gag order" — if she can threaten a punishment of at least $\delta l(\theta)$ for doing so. Clearly, the threat of retaliating against whistleblowing is empty if $p < \delta l(\theta)$. Consequently, the expected outcome must be close enough to x^E for deterrence to be feasible.

There is also a more subtle limitation on the ability to constrain whistleblowing. Even if retaliation for whistleblowing can remove the manager's downside risk from the revelation a bad type, it introduces a problem for the allocation of retaliation effort. Without the possibility of whistleblowing, the manager induces effort optimally by punishing based on type. Since whistleblowing may only occur when $\theta = \underline{\theta}$, the manager cannot punish maximally for *both* type and whistleblowing. Thus punishing the employee for whistleblowing requires the substitution of retaliation away from punishing bad types.

The manager's punishment strategy must therefore take on one of two forms. First, she may continue to punish based on type (of course, if $\overline{p} < \delta l(\theta)$, then she *must* condition on type). Under this strategy, the equilibrium changes from that in Γ^n because the employee can freely whistleblow. Second, she may punish based on whether the employee whistleblows. Since it would only be necessary to threaten a punishment of $\delta l(\theta)$ for whistleblowing, the manager can reserve $\overline{p} - \delta l(\theta)$ for punishing by type. The resulting equilibrium would then be similar to that of Γ^n, except with a reduced punishment capacity.

Reporting and Whistleblowing. E wishes to reveal θ whenever P's posterior beliefs are incorrect in a manner that is disadvantageous to E. For example, if $x^M < \underline{x} < x^E$, then M would ensure that $\mu > \tilde{\mu}$ whenever $\theta = \overline{\theta}$. This allows M to retain decision-making authority in period 2. But M has no incentive to report that $\theta = \underline{\theta}$ if $\mu > \tilde{\mu}$, as this would result in P revoking her authority. Only E would then want to reveal $\underline{\theta}$.

It is evident that E and M's revelation incentives occasionally overlap. For instance, when $x^E > \overline{x}$ and $x^M < \underline{x}$, both E and M wish to reveal that $\theta = \overline{\theta}$ if $\mu_r < \tilde{\mu}$. Since reporting and whistleblowing are costless, there exist equilibria in which either player may report for some game histories. All such equilibria are identical, however, 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 selection rule eliminates many such equilibria, and also preserves the same managerial reporting strategy as that in Γ^n (9), thus maintaining consistency between the two games.

Let $p(\theta)$ denote the anticipated punishment for whistleblowing under type θ, and recall that μ_w is P's posterior belief of $\overline{\theta}$ prior to the revelation of w. At an optimal whistleblowing strategy, E

will report θ if the punishment is not too severe and the managerial report did not convey θ as she would wish:

$$w^* = \begin{cases} \theta & \text{if } p(\theta) < l(\theta), \text{ and either} \\ & \quad x^E > \underline{x}, \theta = \underline{\theta}, \text{ and } \mu_w \geq \tilde{\mu}, \text{ or} \\ & \quad x^E < \overline{x}, \theta = \overline{\theta}, \text{ and } \mu_w < \tilde{\mu} \\ \emptyset & \text{otherwise.} \quad \blacksquare \end{cases} \tag{13}$$

Period 1 Approval. How does the possibility of whistleblowing affect M's initial approval decision? In Γ^n, this decision (10) was sincere because M could reveal through her report that its state-dependent preferences were identical to P's. This intuition is only amplified by whistleblowing, as it reduces M's ability to deceive P when their state-dependent preferences do not agree. Thus M's period 1 approval decision, a_1^*, remains sincere in Γ^w. The proof of this result is virtually identical to that in Γ^n, and is therefore omitted.

These results demonstrate that any consequential differences between Γ^n and Γ^w must lay in the punishment and effort strategies. To examine these, it will be useful to focus the analysis on the two non-trivial cases of the model. In the first, analogous to Proposition 1(i), M is "aggressive" in the sense that $x^M < \underline{x} < x^E$. Here M finds both project types preferable to q, while E and P would prefer the cancellation of the low type project. In the second, analogous to Proposition 1(iii)-(iv), the manager is "conservative," with $x^M > \overline{x} > x^E$. Here M wishes to cancel all projects, while E and P would prefer to approve the high type project. These cases are the only ones in which whistleblowing can have any effect. If E and M were in the same interval, then E would never have an incentive to whistleblow. If M and P were in the same interval, then M would act exactly as P would and neither player's incentives would be affected by whistleblowing.

Case 1 (Aggressive Manager): $x^M < \underline{x} < x^E$. In this case, M approves all projects in period 1, and P can infer θ from x_1 as well as any reporting or whistleblowing decisions. If M's punishment conditions only on type, then as in (8) she penalizes the low type by $\overline{p}$. Since whistleblowing in not deterred, (13) implies that P always has correct posterior beliefs about θ; *i.e.*, $\mu > (<)\ \tilde{\mu}$ whenever $\theta = \overline{\theta}$ ($\underline{\theta}$). As a result, the project is approved in period 2 if and only if it is of the high type. E's objective can therefore be written as:

$$\begin{aligned} U^E &= e(1+\delta)\left[b^E \int_X x f(x|\overline{\theta})dx - k^E\right] + (1-e)\left[b^E \int_X x f(x|\underline{\theta})dx - k^E\right] - (1-e)\overline{p} - ce^2 \\ &= b^E(e\overline{x} + (1-e)\underline{x}) - k^E + \delta e(b^E\overline{x} - k^E) - (1-e)\overline{p} - ce^2. \end{aligned} \tag{14}$$

Straightforward optimization yields the optimum effort level:

$$e_\theta^{w*} = \frac{b^E(\overline{x} - \underline{x}) + \delta(b^E\overline{x} - k^E) + \overline{p}}{2c}. \tag{15}$$

Compared to Proposition 1(i), it is easily seen that $e_\theta^{w*} < e^{n*}$. Punishing by type induces lower effort in a world with whistleblowing because E is assured of the project's cancellation if $\theta = \underline{\theta}$.

By contrast, if $\overline{p} \geq \delta l(\underline{\theta})$, then the punishment can condition on both whistleblowing and type. Under this strategy, M threatens a punishment of $\delta l(\underline{\theta})$ for whistleblowing whenever E would have an incentive to do so; *i.e.*, when $\theta = \underline{\theta}$ and x_1 is high. As with punishing by type, the remainder of her punishment capacity (either $\overline{p} - \delta l(\underline{\theta})$ or $\overline{p}$) is reserved for realizations of $\underline{\theta}$. By (13), this punishment strategy will successfully deter whistleblowing, and the resulting equilibrium is qualitatively similar to that in Γ^n.

As in Γ^n, M's reporting strategy ensures that P always correctly infers the high type, but not the low type. This happens because M only reports when $\theta = \overline{\theta}$ and x_1 is low, and so P must infer θ from x_1 when no report is made. Analogously to Proposition 1, the MLRP (2) implies a cutoff standard $\tilde{x}^w$ below which P infers $\mu < \tilde{\mu}$.[13] E's objective is thus:

$$\begin{aligned} U^E &= e(1+\delta)\left[b^E \int_X x f(x|\overline{\theta})dx - k^E\right] + (1-e)\left[1+\delta(1-F(\tilde{x}^w|\underline{\theta}))\right]\left[b^E \int_X x f(x|\underline{\theta})dx - k^E\right] - \\ &\quad (1-e)\left[(1-F(\tilde{x}^w|\underline{\theta}))(\overline{p}-\delta l(\underline{\theta})) + F(\tilde{x}^w|\underline{\theta})\overline{p}\right] - ce^2 \\ &= (1+\delta)\left[b^E(e\overline{x}+(1-e)\underline{x}) - k^E\right] - \delta(1-e)(b^E\underline{x} - k^E) - (1-e)\overline{p} - ce^2. \end{aligned} \tag{16}$$

Performing the straightforward optimization, the effort induced by this punishment strategy is:

$$e_w^{w*} = \frac{(1+\delta)b^E(\overline{x}-\underline{x}) + \delta(b^E\underline{x} - k^E) + \overline{p}}{2c}. \tag{17}$$

Comparing (15) and (17) reveals that $e_\theta^{w*} = e_w^{w*}$. While objective (16) obviously differs from (14), E's incentives remain unchanged because she is indifferent between whistleblowing and staying silent. Under either punishment scheme, there is effectively a penalty of $\overline{p}$ for a realization of $\underline{\theta}$.

Although M's punishment strategy does not affect effort, its consequences are significant for both P and M. Since more information is revealed when whistleblowing is not deterred, P strictly prefers that M punishes only by type. Likewise, M strictly prefers to punish whistleblowing, as this may allow her to retain her authority when $\theta = \underline{\theta}$.[14]

Figure 2 illustrates the equilibrium revelation of information under the two punishment strategies. The next result formally ties these derivations together to characterize the equilibrium effort and punishment strategies.

[Figure 2]

[13] It can also be shown that $\tilde{x}^w > \tilde{x}^n$, and thus P is less permissive when whistleblowing is possible.

[14] If the strategy of punishing whistleblowing imposed a fixed cost on the M, then M might prefer punishing only types to punishing whistleblowing when m is sufficiently low.

Proposition 2 *Principal's Standard, Employee Effort and Punishment Under an Aggressive Manager. There exists some* $\tilde{x}^w \in X$ *such that* $\mu < (>)\ \tilde{\mu}$ *for* $x_1 < (>)\ \tilde{x}^w$. *E's effort is* $e^{w*} = \frac{b^E(\overline{x}-\underline{x})+\delta(b^E\overline{x}-k^E)+\overline{p}}{2c}$.

If $\overline{p} < \delta l(\underline{\theta})$, *M's punishment is* $p^{w^*}(x_1,\theta,w) = \begin{cases} \overline{p} & \text{if } \theta = \underline{\theta} \\ 0 & \text{if } \theta = \overline{\theta}, \end{cases}$ *and if* $\overline{p} \geq \delta l(\underline{\theta})$,

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

With an aggressive manager, the ability to whistleblow potentially changes each player's strategy. The manager may divert some disciplining authority toward suppressing whistleblowing. The employee's effort is affected by the punishment strategy as well as the lower risk of acquiescing to a low type project in the second period. Finally, the politician may be in a better informational position to assess the manager's decision-making. The next section discusses some of the implications of these strategies.

Case 2 (Conservative Manager): $x^M > \overline{x} > x^E$. In this case, M rejects all first period projects, and therefore wishes to *discourage* effort. If M's punishment conditions only on type, then she penalizes the high type by $\overline{p}$. Since $a_1^* = 0$, Bayes' Rule trivially implies that P's beliefs prior to any reporting are simply $\mu_r = e^*$. If $\mu_r > \tilde{\mu}$, then M has an incentive to report on the low type, while if $\mu_r \leq \tilde{\mu}$, E has an incentive to whistleblow on the high type. Accordingly, either M or E will reveal θ when P's beliefs are incorrect. They both remain silent otherwise. Thus as in the aggressive manager case, P always has correct posterior beliefs about θ.

In period 2, $a_2^* = 1$ if and only if $\theta = \overline{\theta}$. E's objective can therefore be written as:

$$U^E = -e\overline{p} + \delta e(b^E\overline{x} - k^E) - ce^2. \tag{18}$$

Straightforward optimization yields the optimum effort level:

$$e_\theta^{w*} = \max\left\{0, \frac{\delta(b^E\overline{x} - k^E) - \overline{p}}{2c}\right\}. \tag{19}$$

At an interior solution, this is the same expression as in Proposition 1(iii). But unlike the game without whistleblowing, there is no analog to the "corner" case of Proposition 1(iv), because θ is always revealed and P does not have to infer its value imperfectly from its conjecture of e. Thus, punishing by type induces weakly *higher* effort when whistleblowing is possible.

M could also adopt the strategy of punishing whistleblowing if $\overline{p} \geq \delta l(\overline{\theta})$. There are two possible subcases. In the first, $e > \tilde{\mu}$, and P infers from Bayes' Rule that $\mu_r = e$. Under these beliefs, E

never needs to reveal that $\theta = \overline{\theta}$, as P would revoke M's authority in period 2 unless M reveals that $\theta = \underline{\theta}$. Since there is no game history for which E would blow the whistle, M can punish maximally based on type. E's objective is then identical to (18) and so $e_w^{w*} = e_\theta^{w*}$.

The second subcase occurs when the solution to (18) does not satisfy $e \geq \tilde{\mu}$.[15] Now P has pessimistic beliefs ($\mu_r = e < \tilde{\mu}$), and is inclined to retain M's authority in period 2. Moreover, M dissuades E from whistleblowing and would never unilaterally reveal that $\theta = \overline{\theta}$. P then always allows M to cancel the project in period 2. E's objective becomes:

$$U^E = -e(\overline{p} - \delta l(\overline{\theta})) - ce^2. \tag{20}$$

Solving for these two subcases yields the following effort levels:

$$e_w^{w*} = \begin{cases} \frac{\delta(b^E\overline{x} - k^E) - \overline{p}}{2c} & \text{if } \frac{\delta(b^E\overline{x} - k^E) - \overline{p}}{2c} \geq \tilde{\mu} \quad (i) \\ 0 & \text{otherwise.} \qquad\qquad\quad (ii) \end{cases} \tag{21}$$

As intuition might suggest, under the strategy of deterring whistleblowing, the game becomes almost identical to Γ^n. As in Proposition 1, the intuition of part (ii) of (21) is that a low but positive effort level cannot be sustained in equilibrium. Since M cancels the project in period 2, E would receive a negative payoff by choosing $e > 0$, and therefore does strictly better with the corner effort level of zero. Note also that P continues to prefer that M punish only by type, since this strategy can yield a higher effort than punishing whistleblowing.

In comparing the two punishment strategies, it is clear that M weakly prefers e_w^{w*} to e_θ^{w*}. The best outcome for the conservative manager is part (ii) of (21). Since there is no possibility of the high type, there is also no possibility that P would revoke M's authority. However, this outcome is impossible if P's beliefs about $\overline{\theta}$ are so strong that she removes control from M even without whistleblowing. In this environment, P's posterior beliefs will always be correct, and M is left to punish only by type. Thus a conservative manager will punish whistleblowing whenever possible, and by type otherwise. The following result formalizes this argument to establish equilibrium punishment and effort strategies.

Proposition 3 *Effort and Punishment Under a Conservative Manager. If $\overline{p} < \delta l(\overline{\theta})$, then E's effort is $e^{w*} = e_\theta^{w*}$ and M's punishment is* $p^{w^*}(x_1, \theta, w) = \begin{cases} \overline{p} & \text{if } \theta = \overline{\theta} \\ 0 & \text{otherwise.} \end{cases}$

If $\overline{p} \geq \delta l(\overline{\theta})$, then $e^{w*} = e_w^{w*}$ and:

$$p^{w^*}(x_1, \theta, w) = \begin{cases} \overline{p} & \text{if } \theta = \overline{\theta} \text{ and } \frac{\delta(b^E\overline{x} - k^E) - \overline{p}}{2c} \geq \tilde{\mu}, \text{ or} \\ & \quad \theta = \overline{\theta}, \frac{\delta(b^E\overline{x} - k^E) - \overline{p}}{2c} < \tilde{\mu}, \text{ and } w = \theta \\ \overline{p} - \delta l(\overline{\theta}) & \text{if } \theta = \overline{\theta}, \frac{\delta(b^E\overline{x} - k^E) - \overline{p}}{2c} < \tilde{\mu}, \text{ and } w = \emptyset \\ 0 & \text{if } \theta = \underline{\theta}. \quad \blacksquare \end{cases}$$

[15] As in Γ^n, this argument uses the refinement that selects the equilibrium with the highest effort level by E.

The next figure illustrates the equilibrium revelation of information with a conservative manager.

[Figure 3]

Both cases of Γ^w yield several insights into the incentives surrounding 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 cannot occur, the politician benefits from whistleblowing *given* a certain level of employee effort. However, the *ex ante* effects on employee effort may not benefit the politician. The next section examines several implications of these results for whistleblower policy.

It is worth noting finally that because the model focuses primarily on *ex ante* organizational incentives, whistleblowing affects the manager only through her allocation of employee incentives. In equilibrium, it neither "disciplines" her to report θ more often, nor changes her (sincere) approval decisions. This happens in part because reports fully reveal θ, and the manager's preferences are common knowledge. Relaxing these assumptions, or introducing moral hazard problems on the part of the manager, might induce more managerial strategic behavior.

4. Whistleblower Policy

4.1 Allowing Whistleblowing

One simple way to assess the impact of basic whistleblower protections is by comparing the effort levels predicted by Γ^n and Γ^w. The following result establishes that whistleblowing moves effort in the opposite direction from that which the manager would prefer. This helps the principal when the manager is conservative, but not when she is aggressive. In some cases, the principal may even prefer to disallow whistleblowing with an aggressive manager.

Comment 1 *Whistleblowing and Effort. (i) Under an aggressive manager,* $e^{w*} \leq e^{n*}$. *Under a conservative manager,* $e^{w*} \geq e^{n*}$.

(ii) Under an aggressive manager, if $\overline{p} \geq \delta l(\underline{\theta})$ *then P's expected utility is higher in* Γ^n *than in* Γ^w. *Under a conservative manager, P's expected utility is higher in* Γ^w *than in* Γ^n. ■

When M is aggressive, the ability to whistleblow reduces effort in two ways. If M punishes only types, then E faces lower downside risk from a realization of $\underline{\theta}$. Because of whistleblowing, this type results in a zero payoff in Γ^w, versus a negative expected payoff in Γ^n. If M also punishes whistleblowing, then M disciplines E less when $\theta = \underline{\theta}$ in order to reserve sufficient punishment capacity to deter whistleblowing. Thus when M can punish whistleblowing, the combination of low effort and no whistleblowing lowers P's payoff relative to Γ^n.

When M is conservative, effort levels are weakly higher than in Γ^n, and can be strictly higher when M punishes only by type. This happens because whistleblowing allows E to reveal $\overline{\theta}$ even if her effort level is low. Unlike the aggressive manager case, effort does not increase due to the diversion of discipline, because E effectively "cancels" the project by choosing $e = 0$ in both games. The combination of whistleblowing and higher effort results in a higher payoff for P than in Γ^n.

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 $\delta = 0.9$ and $c = 1$. The policy space is $X = [0, 1]$. The outcomes are distributed uniformly for type $\underline{\theta}$; $f(x_t|\underline{\theta}) = 1$, while the density is linear for type $\overline{\theta}$; $f(x_t|\overline{\theta}) = 2x_t$. It is straightforward to calculate that $\underline{x} = 0.5$, $\overline{x} = 0.667$, and $\tilde{\mu} = 0.333$.

The table below compares the payoffs across Γ^w and Γ^n for $\overline{p} = 0.16, 0.2$. These values of $\overline{p}$ were chosen because $0.16 < \delta l(\underline{\theta}) < 0.2$. This implies that M can only feasibly punish by type when $\overline{p} = 0.16$, but can also choose to whistleblowing when $\overline{p} = 0.2$.

Table 1
Examples of Managerial Strategies

Game	Γ^w			Γ^n	
$\overline{p}$	0.16	0.2		0.16	0.2
Punishment Strategy	Type	Type*	Whis.	Type	Type
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					

When $\overline{p} = 0.2$, M's expected payoff is higher when she deters whistleblowing, even though E's effort is the same under either punishment strategy. This is because the cutoff standard for inferring that $\mu = \tilde{\mu}$ is $\tilde{x}^w = 0.699$, which gives M an over 30% chance of retaining managerial control even when the project type is low. P would prefer that M punish only by type, since this would effectively always reveal θ in period 2. Given that M does not follow that strategy, P would do better in Γ^n. 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 E reduces equilibrium effort. M can only punish by type in this environment, and so E blows the whistle whenever P wrongly infers that $\theta = \overline{\theta}$. P therefore always makes the correct assignment of decision-making rights in period

2. As Comment 1 predicts, effort is lower in Γ^w than in Γ^n because E faces a lower downside from a low type. However, the informational gains from whistleblowing more than offset the lost effort. Thus the ability to whistleblow helps the politician.

4.2 Optimal Punishments

Given that some level of whistleblower protections are inevitable in practice, what extent of legal protection would principals favor? Laws such as the WPA typically provide a set of procedural guarantees to facilitate employee reporting.[16] They also contain provisions, such as promises of confidentiality and injunctions against managerial actions, that make retaliation against whistleblowers more difficult. Perhaps most significantly, such laws prohibit explicit retaliation at all. 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.
>
> (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.

One effect of such laws is to discourage managers from punishing whistleblowing. The analysis of Section 3.2 establishes that a switch in strategy to punishing only types or performance will benefit the politician. Thus, politicians would stand to benefit greatly if whistleblowing laws have the effect of regulating *only* the kinds of strategies employed by managers.

The model developed here suggests that restricting merely managerial strategies would be difficult to accomplish. While whistleblower protection laws clearly make retaliation against employees more difficult, they can also change employee effort and whistleblowing incentives. In particular, an employee could invoke whistleblower protection to reduce the scope of *all* managerial disci-

[16]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.

plinary action.[17] This argument assumes, as the model does, that whistleblower complaints could be initiated at low cost.

To formalize this idea, suppose that the maximum legal punishment against a whistleblower were reduced to $\overline{p}^w < \overline{p}$, while leaving the maximum punishment for non-whistleblowers at $\overline{p}$. If M wishes to deter whistleblowing for some θ and $\overline{p}^w < \delta l(\theta)$, then M obviously has no choice but to adopt a strategy of punishing types. But there is an additional problem for M. Even if $\overline{p}^w \geq \delta l(\theta)$, M cannot deliver a punishment of $\overline{p}$ if E blows the whistle. E can then circumvent *any* sufficiently high punishment simply by invoking whistleblowing protections (*i.e.*, choosing $w = \theta$). As the result shows, in both the aggressive and conservative manager cases examined in Section 3, no punishment of more than $\overline{p}^w$ need be part of an optimal punishment strategy.

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

Limiting only whistleblower retaliation to $\overline{p}^w$ effectively constrains all punishments to be no greater than $\overline{p}^w$. Thus, any such limitation is effectively a reduction of $\overline{p}$. The specific protection of whistleblowing therefore has effects similar to those of limiting managerial latitude more generally. Analogously to Comment 1, the next comment characterizes the comparative statics on $\overline{p}^w$ or $\overline{p}$: stronger whistleblower protections are often harmful to the principal under ambitious managers, but helpful under conservative managers.

Comment 3 *Optimal Whistleblower Protection. (i) Under an aggressive manager, e^{w*} is increasing in $\overline{p}$. Under a conservative manager, e^{w*} is weakly decreasing in $\overline{p}$.*

(ii) Under an aggressive manager, P's expected utility is weakly increasing in $\overline{p}$ except at $\overline{p} = \delta l(\underline{\theta})$. Under a conservative manager, P's expected utility is weakly decreasing in $\overline{p}$. ■

The intuition of part (i) is that reducing $\overline{p}$ will also reduce M's ability to induce performance in her desired direction. Since an aggressive manager desires higher effort, reducing $\overline{p}$ lowers both E's effort and P's utility at the margin. A conservative manager reverses this logic. Since a conservative manager wants lower effort, equilibrium effort increases with whistleblower protection.

While part (i) suggests that stronger whistleblower protections only benefit political principals when a manager is conservative, part (ii) raises one important exception. As Table 1 illustrated, a

[17] 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).

manager can punish only by type when $\overline{p} < \delta l(\underline{\theta})$, as deterring whistleblowing is infeasible. As a result, in the neighborhood near $\delta l(\underline{\theta})$, the reduction in effort from a lower value of $\overline{p}$ is compensated for by the certainty of revealing a low type project. Whistleblower protections therefore may help the principal in an environment in which whistleblowing is routinely suppressed.[18]

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 depend greatly on the preferences of the manager relative to the status quo.

4.3 Explicit Rewards

A centerpiece of American whistleblowing legislation is the False Claims Act, 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 the politician to reclaim damages more easily, though this aspect will not be addressed here.

To explore this feature, suppose that E receives from M a proportion $\pi \in (0, 1)$ of M's surplus when θ is revealed and a_1 is not the decision that P would have made. The surplus is simply the expected difference between the payoff from M's period 1 action and P's preferred action, or $|b^M E[x_1|\theta] - k^M|$. For example, given a cutoff standard $\tilde{x}^\pi$, E receives the reward if $\theta = \underline{\theta}$ and $x_1 > \tilde{x}^\pi$. Importantly, I assume that there is no distinction between whether M or E reports θ, so that M cannot benefit from pre-emptively revealing damaging information. This assumption simplifies matters by ensuring that M's period 1 approval decision remains "sincere," just as in Γ^w. As a result, much of the extended model can be analyzed simply by examining the effect of rewards on E's objective. For example, if M is ambitious and punishes by type, then E's objective is identical to (14), with the exception of an additional "reward" term.

One immediate effect of a higher π is to reduce the set of employees for which punishing whistleblowing is possible, since a whistleblower expects to lose less than $\overline{p}$ by making a report. As the main model establishes, this generally benefits the principal. The effect on effort depends on managerial preferences. The following result establishes that raising π is similar to reducing $\overline{p}$; thus, *qui tam* rewards may not increase employee effort.

[18] 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 only by type when m is low (*i.e.*, M is less career-minded or more policy-minded), and would punish whistleblowing if feasible when m is high (see footnote 14). 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 of whistleblower protections in an agency will increase the proportion of political appointees.

Comment 4 Qui Tam *Incentives. Under an aggressive manager, e^{w*} is weakly decreasing in π. Under a conservative manager, e^{w*} is weakly increasing in π.* ■

The intuition of the aggressive manager case is straightforward: managerial "wrongdoing" occurs only when $\theta = \underline{\theta}$. Thus *qui tam* provisions actually increase the payoff from type $\underline{\theta}$, and therefore encourage lower effort. With a conservative manager, the logic is reversed: M rejects the high type project, and so *qui tam* provisions encourage the employee to exert more effort. Thus as with other whistleblower laws, the False Claims Act establishes effective *ex post* incentives that may conflict with optimal *ex ante* incentives.

This result may help to explain some of the variation observed in *qui tam* laws. The percentage of damages which *qui tam* relators could claim has varied considerably over history. The original 1863 False Claims Act allowed up to 50%. Amendments to the law in 1943 addressed what Congress believed to be "parasitic" whistleblowers, and decreased the relator's share to 10%-25%, depending on the type of case. The limits were raised by 1986 amendments to 15%-30%, and generated a large increase in recovered funds.[19] Comment 4 predicts that the proportion π should vary with the preferences of managers in the bureaucracy.

5. Conclusions

For well over a century, policy-makers have recognized the importance of whistleblowers in aiding the transmission of information from 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.

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 of the model suggest that whistleblower protections do very well on the latter, but relatively poorly on the former. In fact, under some circumstance it may even be optimal for the politician not to allow whistleblowing.

A key variable that emerges from the model is the manager's preferences relative to those of the politician. An aggressive manager, who is more inclined than the politician to approve a project, will typically use her ability to punish employees in ways that increase their effort. In this case, common whistleblower protections can reduce the power of employee incentives. By

[19]The 1986 amendments were followed by WPA provisions which allowed plaintiffs to take some cases directly to court. The recovered amounts increased from less than $10 million to over $100 million per year.

contrast, conservative managers wish to suppress effort, and so whistleblower protections will have the salutary effect (from the politician's perspective) of increasing employee effort.

It is finally worth considering how variation in managerial preferences may be measured empirically. One way is simply to examine the composition of agency personnel relative to that of their political principals (*e.g.*, Lewis 2004). 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. Another is to link managerial aggressiveness with organizational structure. In the model, an aggressive manager is one who places more emphasis on avoiding Type II error (relative to the "default" of $x_t = q$), while a conservative manager is more concerned with avoiding Type I error. An extensive literature examines the impact of organizational design on Type I and II errors (Bendor 1985, Heimann 1993, 1997, Carpenter and Ting 2006). Using these theories, it should be possible to derive optimal whistleblowing policies as a function of organizational design.

APPENDIX

Lemma 1 *Punishment without Whistleblowing.* $p^{n*}(x_1,\theta) = \begin{cases} \overline{p} & \text{if } \theta = \underline{\theta} \text{ and } x^M < \overline{x}, \\ & \quad \text{or } \theta = \overline{\theta} \text{ and } x^M > \overline{x} \\ 0 & \text{otherwise.} \end{cases}$ ■

Proof. Throughout, let $g^i(\theta)$ denote player i's expected equilibrium payoff given type θ, excluding any punishments.

It is first necessary to determine which incentives M wishes to provide. Given some effort level e, M's objective can be written in the following general form:

$$U^M = eg^M(\overline{\theta}) + (1-e)g^M(\underline{\theta}). \tag{22}$$

It is clear by (22) that M would provide incentives for higher effort if $g^M(\overline{\theta}) \geq g^M(\underline{\theta})$, and lower effort otherwise.

There are three cases, depending on the location of x^M. All use Lemmas 2 and 3, but those results do not depend on this lemma.

(i) $x^M > \overline{x}$. By Lemma 3, M rejects all projects at $t = 1$. Bayes' Rule then trivially implies that $\mu = e$. By Lemma 2, $\mu < \tilde{\mu}$ whenever $\theta = \underline{\theta}$. By (7), $s^* = M$ if and only if $\mu < \tilde{\mu}$. Thus if $\theta = \underline{\theta}$, M receives $(1+\delta)m$. If $\theta = \overline{\theta}$, then M receives $(1+\delta)m$ if $e \leq \tilde{\mu}$, and m otherwise. Hence $g^M(\overline{\theta}) \leq g^M(\underline{\theta})$, so M weakly prefers lower effort.

Now consider the incentives that P can provide through the punishment, p. Since $x_1 = q$, p can only condition on θ. Any punishment strategy is thus a pair (p', p''), where $p' \geq 0$, $p'' \geq 0$, and $p' + p'' \leq \overline{p}$. E's objective is then:

$$U^E = eg^E(\overline{\theta}) + (1-e)g^E(\underline{\theta}) - ep' - (1-e)p'' - ce^2.$$

This objective is concave. Clearly, $\arg\max U^E$ is minimized if $\frac{dU^E}{de}$ is minimized, and $\frac{dU^E}{de}$ is minimized if $p' - p''$ is maximized. Thus the optimal punishment strategy is $p^*(x_1, \overline{\theta}) = \overline{p}$ and $p^*(x_1, \underline{\theta}) = 0$.

(ii) $x^M < \underline{x}$. By Lemma 3, M approves all projects at $t = 1$. By (7) and Lemma 2, $\mu > \tilde{\mu}$ and $s^* = M$ whenever $\theta = \overline{\theta}$. Thus if $\theta = \overline{\theta}$, then M receives $(1+\delta)[b^M\overline{x} - k^M + 2m]$. If $\theta = \underline{\theta}$, then M receives at most $(1+\delta)[b^M\underline{x} - k^M + 2m]$. Hence $g^M(\overline{\theta}) > g^M(\underline{\theta})$, so M prefers higher effort.

Since P can now condition on both x_1 and θ, the punishment strategy may be written as $p(x_1|\theta)$, where $p(x_1|\theta) \leq \overline{p}$ for all x_1, θ. E's objective may then be written as:

$$U^E = eg^E(\overline{\theta}) + (1-e)g^E(\underline{\theta}) - e\int_X p(x|\overline{\theta})f(x|\overline{\theta})dx - (1-e)\int_X p(x|\underline{\theta})f(x|\underline{\theta})dx - ce^2. \tag{23}$$

This objective is concave. Analogously to case (i), $\arg\max U^E$ is maximized when $\frac{dU^E}{de}$ is maximized. Differentiating (23) reveals that $\frac{dU^E_\theta}{de}$ is maximized when $\int_X p(x|\underline{\theta})f(x|\underline{\theta}) - p(x|\overline{\theta})f(x|\overline{\theta})dx$ is maximized. Hence the optimal punishment strategy is $p^*(x_1, \overline{\theta}) = 0$ and $p^*(x_1, \underline{\theta}) = \overline{p}$.

(iii) $x^M \in (\underline{x}, \overline{x})$. In this case, Lemma 3 implies that P is fully informed of θ: $\mu = 1$ (0) if $\theta = \overline{\theta}$ ($\underline{\theta}$). M receives $2m$ if $\theta = \underline{\theta}$, and $(1+\delta)[b^M\overline{x} - k^M + 2m]$ if $\theta = \overline{\theta}$. Hence $g^M(\overline{\theta}) > g^M(\underline{\theta})$, so M prefers higher effort. The proof is a straightforward combination of cases (i) and (ii), and is therefore omitted. ■

Lemma 2 *Reporting without Whistleblowing.* $r^* = \begin{cases} \theta & \text{if } \theta = \overline{\theta}, \mu_r < \tilde{\mu}, \text{and } x^M < \underline{x}, \text{ or} \\ & \quad \theta = \underline{\theta}, \mu_r > \tilde{\mu}, \text{and } x^M > \overline{x} \\ \emptyset & \text{otherwise.} \end{cases}$ ■

Proof. Note that M's payoff is maximized by $s = M$, which yields a payoff of at least $m > 0$. There are three subcases, depending on the location of x^M. First, if $x^M < \underline{x}$, then (7) implies that $s^* = M$ iff $\mu \geq \tilde{\mu}$. Suppose that $\mu_r < \tilde{\mu}$. If $\theta = \overline{\theta}$, then it is easily verified that under any optimal reporting strategy, $r^* = \theta$, and hence $\mu = 1$. The minimum reporting rule is therefore uniquely satisfied by: $r^* = \theta$ iff $\theta = \overline{\theta}$, which results in $\mu = 1$ (0) if $r = \theta$ ($= \emptyset$). To verify that this is an equilibrium strategy, note that type $\overline{\theta}$ 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. If $\mu_r \geq \tilde{\mu}$, then the minimum reporting rule would be uniquely satisfied by $r^* = \emptyset$ for all θ. To verify that this is an equilibrium strategy, note simply that $\mu = \tilde{\mu}$ and P's response is then $s = M$, which yields M's maximal payoff. 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. Third, if $x^M \in (\underline{x}, \overline{x})$, then since Lemma 3 implies that $a_1 = 0$ ($= 1$) iff $\theta = \underline{\theta}$ ($= \overline{\theta}$), M need not issue a report; thus $r^* = \emptyset$ for all θ. The resulting strategy is identical to that in Lemma 2. ■

Lemma 3 *Managerial Approval.* $a_1^* = \begin{cases} 1 & \text{if } b^M E[x_1|\theta] - k^M \geq 0 \\ 0 & \text{otherwise.} \end{cases}$ ■

Proof. Consider the case where $x^M < \underline{x}$, so that M would myopically approve both project types. If $\theta = \overline{\theta}$ and M approves the project at $t = 1$, then according to M's reporting strategy (Lemma 2), $\mu \geq \tilde{\mu}$. Thus, by (7), 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 $\delta(b^M\overline{x} - k^M) + (1 + \delta)m$. Thus she does strictly worse by choosing $a_1 = 0$, and so $a_1^*(\overline{\theta}) = 1$ in any equilibrium.

Now consider whether M would ever choose $a_1 = 0$ when $\theta = \underline{\theta}$. Given that $a_1^*(\overline{\theta}) = 1$, P infers $\mu = 0$ if $a_1 = 0$ in any equilibrium, and by (7), 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 cases where $x^M > \underline{x}$ are proven identically and are omitted. ■

Proof of Proposition 1. I begin by solving for an interior effort level, e^n. There are three cases.

Case 1: $x^M < \underline{x}$. By Lemma 3 M approves all project types. Given $\tilde{x}^n$ and the strategies in Lemmas 1 and 2, E's objective is:

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

This objective is clearly concave. Differentiating yields the following first-order condition:

$$(1+\delta)b^E(\overline{x} - \underline{x}) + \delta F(\tilde{x}^n|\underline{\theta})(b^E\underline{x} - k^E) + \overline{p} - 2ce = 0.$$

Solving yields the optimum effort level:

$$e^n = \frac{(1+\delta)b^E(\overline{x} - \underline{x}) + \delta F(\tilde{x}^n|\underline{\theta})(b^E\underline{x} - k^E) + \overline{p}}{2c}. \tag{24}$$

Case 2: $x^M \in (\underline{x}, \overline{x})$. M's preferences now coincide with P's. At $t = 1$, the project is approved if and only if $\theta = \overline{\theta}$. E's objective is:

$$U^E = e(1+\delta)(b^E\overline{x} - k^E) - (1-e)\overline{p} - ce^2.$$

This objective is concave. Solving as before yields:

$$e^n = \frac{(1+\delta)(b^E\overline{x} - k^E) + \overline{p}}{2c}. \tag{25}$$

Case 3: $x^M > \overline{x}$. M rejects all projects at $t = 1$. Note that because $a_1^* = 0$, Bayes' Rule implies $\mu = e^*$. There are two possible solutions. If $e^* > \tilde{\mu}$, then $s^* = P$, $a_2^* = 1$, and E's objective is:

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

This objective is concave, so differentiating and solving produces:

$$e^n = \frac{\delta(b^E\overline{x} - k^E) - \overline{p}}{2c}. \tag{26}$$

Likewise, if $e^* < \tilde{\mu}$, then $s^* = M$, $a_2^* = 0$, and E's objective is $-e\overline{p} - ce^2$, which is also concave. Differentiating and solving produces a solution at $e^n = -p/(2c)$. This value is clearly negative. This

solution trivially satisfies $e^* < \tilde{\mu}$. However, if (26) satisfies $e^n > \tilde{\mu}$, then there are two equilibrium solutions. Since the equilibrium that maximizes E's effort is selected, e^n is given by (26) iff $e^n > \tilde{\mu}$.

By the assumptions on c, $e^n < 1$. By the concavity of all objective functions, the optimal effort level is then $e^{n*} = \max\{0, e^n\}$.

To complete the equilibrium, it is finally necessary to determine the standard $\tilde{x}^n$ at which P believes that $\mu = \tilde{\mu}$. Let $e^{n*}(x_1)$ be the optimal effort implied by a cutoff at x_1, and let $\mu(x_1)$ be the associated posterior belief. Applying (11), $\mu(x_1) = \tilde{\mu}$ if:

$$\frac{f(x_1|\overline{\theta})e^{n*}(x_1)}{f(x_1|\overline{\theta})e^{n*}(x_1) + f(x_1|\underline{\theta})(1 - e^{n*}(x_1))} = \frac{k^P/b^P - \underline{x}}{\overline{x} - \underline{x}}. \tag{27}$$

By (4), $\tilde{\mu} > 0$. Then by (3), it is clear that $\lim_{x_1 \downarrow \min X} \mu(x_1) < \tilde{\mu}$. Let $\tilde{x}^n = \min\{x_1 \mid (27) \text{ holds}\}$ if that set is non-empty, and $\tilde{x}^n = \max X$ otherwise. Then given effort $e^{n*}(\tilde{x}^n)$, MLRP implies that μ is increasing in x_1, and thus $\mu < (>)\ \tilde{\mu}$ for $x_1 < (>)\ \tilde{x}$. ■

Lemma 4 *Reporting and Whistleblowing. In Γ^w, r^* remains as in Lemma 2, and*

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

Proof. Consider first the whistleblowing decision, w. Clearly, if $p(\theta) \geq l(\theta)$, then $w^* = \emptyset$. If $p(\theta) < l(\theta)$, then there are three subcases. First, if $x^E > \overline{x}$, then E's period 2 payoff is maximized by $a_2 = 0$. By (5) and (7), $a_2^* = 0$ 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 $\mu < \tilde{\mu}$. If $\mu_w \geq \tilde{\mu}$ and $\theta = \underline{\theta}$, then E must choose $w^* = \theta$ to ensure that $\mu = 0 < \tilde{\mu}$. Given this strategy, Bayes' rule implies that $\mu \geq \tilde{\mu}$ when $\mu_w \geq \tilde{\mu}$ and $w = \emptyset$ under any whistleblowing strategy. Thus E prefers $w = \theta$ when $\theta = \underline{\theta}$, and by the minimum reporting rule, $w^* = \emptyset$ if $\theta = \overline{\theta}$ 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 $\mu_w < \tilde{\mu}$. Third, $x^E \in (\underline{x}, \overline{x})$, then the arguments of first two cases can be combined straightforwardly to show that $w^* = \theta$ iff $\theta = \underline{\theta}$ $(= \overline{\theta})$ and $\mu_w \geq \tilde{\mu}$ $(< \tilde{\mu})$.

For the reporting decision r, note that M's payoff is maximized by $s = M$, which yields a payoff of at least $m > 0$. There are three subcases, depending on the location of x^M. First, let $x^M < \underline{x}$, which by (7) implies that $s^* = M$ iff $\mu \geq \tilde{\mu}$. Suppose that $\mu_r < \tilde{\mu}$. If $\theta = \overline{\theta}$, then under any optimal reporting strategy, either $r^* = \theta$ or $w^* = \theta$ (resulting in $s^* = M$). The minimum reporting rule is therefore uniquely satisfied by: $r^* = \theta$ iff $\theta = \overline{\theta}$, which results in $\mu = 1$ (0) if $r = \theta$ $(= \emptyset)$. To verify that this is an equilibrium strategy, note that type $\overline{\theta}$ 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. If $\mu_r \geq \tilde{\mu}$, then the minimum reporting rule would be 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. Third, if $x^M \in (\underline{x}, \overline{x})$, then since $a_1^* = 0$ $(= 1)$ iff $\theta = \underline{\theta}$ $(= \overline{\theta})$, M need not issue a report; thus $r^* = \emptyset$ for all θ. The resulting strategy is identical to that in Lemma 2. ∎

Proof of Proposition 2. The existence of $\tilde{x}^w$ is proved identically to that of $\tilde{x}^n$ in Proposition 1.

I next establish that one of the two punishment strategies must be optimal. If p depends only on x_1 and θ, then by the argument in part (ii) of Lemma 1, the optimal punishment is:

$$p(x_1, \theta, w) = \begin{cases} \overline{p} & \text{if } \theta = \underline{\theta} \\ 0 & \text{if } \theta = \overline{\theta}. \end{cases} \tag{28}$$

Now suppose that M additionally conditions on w. By Lemma 4, $\mu_w \geq \tilde{\mu}$ if $\theta = \overline{\theta}$, and $\mu_w \geq \tilde{\mu}$ when $\theta = \underline{\theta}$ iff $x_1 \geq \tilde{x}^w$. Thus any $p(x_1, \theta, w) > 0$ can benefit M only if $\theta = \underline{\theta}$, $x_1 \geq \tilde{x}^w$, and. By (13), whistleblowing yields E an expected utility change of $\delta l(\theta)$. Thus to deter whistleblowing, the punishment must satisfy $p(x_1, \underline{\theta}, \theta) \geq \delta l(\underline{\theta})$. Clearly, no punishment exceeding $\delta l(\underline{\theta})$ is necessary to deter whistleblowing. Additionally, Lemma 4, (5), and (7) imply that if punishing by $\delta l(\underline{\theta})$ is optimal for some $x_1' \geq \tilde{x}^w$ then punishing by $\delta l(\underline{\theta})$ must be optimal for all $x_1 \geq \tilde{x}^w$. The optimal punishment for whistleblowing must then be $\delta l(\underline{\theta})$ for $\theta = \underline{\theta}$, $w = \theta$, and $x_1 \geq \tilde{x}^w$.

Again applying the argument of part (ii) of Lemma 1, M also punishes the realization of type $\underline{\theta}$ by the maximum possible amount. Thus if $\theta = \underline{\theta}$ and $x_1 \geq \tilde{x}^w$, M punishes by $\overline{p} - \delta l(\underline{\theta})$, and if $\theta = \underline{\theta}$ and $x_1 < \tilde{x}^w$, M punishes by $\overline{p}$. Combining these arguments, the optimal punishment strategy that also conditions on w is:

$$p(x_1, \theta, w) = \begin{cases} \overline{p} & \text{if } \theta = \underline{\theta} \text{ and either } x_1 < \tilde{x}^w \text{ or } w = \theta \\ \overline{p} - \delta l(\underline{\theta}) & \text{if } \theta = \underline{\theta}, x_1 \geq \tilde{x}^w, \text{and } w = \emptyset \\ 0 & \text{if } \theta = \overline{\theta}. \end{cases} \tag{29}$$

Both punishment strategies induce the same effort level e^{w*}, derived in (15) and (17).

To see which punishment strategy M adopts in equilibrium, note first that if $\overline{p} < \delta l(\underline{\theta})$, then (29) is infeasible and M uses (28). Otherwise, if $\overline{p} \geq \delta l(\underline{\theta})$, then since a_1^* identical under both strategies, it is sufficient to compare the two strategies' period 2 payoffs. The punishment strategy (29) yields a higher expected payoff for M if:

$$\begin{aligned} e^{w*}(b^M \overline{x} - k^M + m) + (1 - e^{w*})[(1 - F(\tilde{x}^w))(b^M \underline{x} - k^M + m)] &\geq e^{w*}(b^M \overline{x} - k^M + m) \\ \Leftrightarrow (1 - e_w^{w*})(1 - F(\tilde{x}^w))(b^M \underline{x} - k^M + m) &\geq 0. \end{aligned}$$

Observe that since M is aggressive, the left-hand side of the last expression is always non-negative, thus establishing the result. ■

Proof of Proposition 3. By an argument symmetric with that in Proposition 2 (using the argument in part (i) of Lemma 1), the optimal punishment that conditions only on x_1 and θ is:

$$p(x_1, \theta, w) = \begin{cases} \overline{p} & \text{if } \theta = \overline{\theta} \\ 0 & \text{if } \theta = \underline{\theta}. \end{cases} \tag{30}$$

To derive the optimal punishment strategy that also conditions on w, note that if $p(x_1, \overline{\theta}, \theta) < \delta l(\overline{\theta})$, then E's best response is $w = \theta$ when $\theta = \overline{\theta}$. Applying the argument in Proposition 2 yields:

$$p(x_1, \theta, w) = \begin{cases} \overline{p} & \text{if } \theta = \overline{\theta} \text{ and } w = \theta \\ \overline{p} - \delta l(\overline{\theta}) & \text{if } \theta = \overline{\theta} \text{ and } w = \emptyset \\ 0 & \text{if } \theta = \underline{\theta}. \end{cases} \tag{31}$$

To see which punishment strategy M adopts in equilibrium, note that (30) and (31) induce effort levels e_θ^{w*} and e_w^{w*}, respectively. If $\overline{p} < \delta l(\overline{\theta})$, then clearly (31) is infeasible and M uses (30).

Otherwise, if $\overline{p} \geq \delta l(\overline{\theta})$, there are two cases. First, if $\frac{\delta(b^E\overline{x} - k^E) - \overline{p}}{2c} \geq \tilde{\mu}$, then $e_w^{w*} = e_\theta^{w*}$. If M punishes whistleblowing, then since $x_1 = q$ and $\mu_r = e_w^{w*}$, in equilibrium $\mu > \tilde{\mu}$ unless M reports θ. By Lemma 4, this occurs when $\theta = \underline{\theta}$, and so $\mu > \tilde{\mu}$ $(< \tilde{\mu})$ when $\theta = \overline{\theta}$ $(= \underline{\theta})$. Likewise, by Lemma 4, when M conditions only on θ, $\mu > \tilde{\mu}$ $(< \tilde{\mu})$ when $\theta = \overline{\theta}$ $(= \underline{\theta})$. Thus under both punishment strategies, θ is revealed and P's responses are identical. Therefore $e^{w*} = e_w^{w*}$, and so $p^*(x_1, \overline{\theta}, w) = \overline{p}$ and $p^*(x_1, \underline{\theta}, w) = 0$.

Second, if $\frac{\delta(b^E\overline{x} - k^E) - \overline{p}}{2c} < \tilde{\mu}$, then $e_w^{w*} = 0 \leq e_\theta^{w*}$. Thus by punishing whistleblowing, P receives a second period payoff of m, while if $e_\theta^{w*} > 0$ then P expects strictly less than m. M therefore weakly prefers the strategy (31) of punishing whistleblowing, and so $p^*(x_1, \overline{\theta}, \theta) = \overline{p}$, $p^*(x_1, \overline{\theta}, \emptyset) = \overline{p} - \delta l(\overline{\theta})$, and $p^*(x_1, \underline{\theta}, w) = 0$. ■

Proof of Comment 1. (i) Consider first the aggressive manager case. To show that $e^{w*} \leq e^{n*}$, by (24) and (15) it is sufficient to show:

$$\begin{aligned} b^E(\overline{x} - \underline{x}) + \delta(b^E\overline{x} - k^E) + \overline{p} &\leq (1+\delta)b^E(\overline{x} - \underline{x}) + \delta F(\tilde{x}^n|\underline{\theta})(b^E\underline{x} - k^E) + \overline{p} \\ \Leftrightarrow (1 - F(\tilde{x}^n|\underline{\theta}))(b^E\overline{x} - k^E) &\leq b^E(\overline{x} - \underline{x}). \end{aligned}$$

This expression holds if $(1 - F(\tilde{x}^n|\underline{\theta}))(b^E\underline{x} - k^E) < 0$, which follows from the fact that $b^E\underline{x} - k^E < 0$ when M is aggressive.

For a conservative manager, if $\overline{p} < \delta l(\overline{\theta})$, then M punishes only by type and the result follows immediately from Proposition 1(iii)-(iv) and (19), with the inequality strict for $\frac{\delta(b^E\overline{x} - k^E) - \overline{p}}{2c} < \tilde{\mu}$.

If $\overline{p} \geq \delta l(\overline{\theta})$, then M punishes by whistleblowing and type. By Propositions 1(iii)-(iv) and 3, this implies that $e^{n*} = e^{w*}$, thus establishing the result.

(ii) If M is aggressive and $\overline{p} \geq \delta l(\underline{\theta})$, then let U^{Pg} denote P's equilibrium expected utility under game Γ^g ($g \in \{w, n\}$). P's expected utility is then $U^{Pg} = e^{g*}(1+\delta)(b^P\overline{x} - k^P) + (1-e^{g*})[1+\delta(1-F(\tilde{x}^g|\underline{\theta}))](b^P\underline{x} - k^P)$. Suppose to the contrary that $U^{Pw} > U^{Pn}$. By part (i), $e^{w*} \leq e^{n*}$. Then by using standard $\tilde{x}^w$ in Γ^n, P receives $U^{Pn'} = e^{n*}(1+\delta)(b^P\overline{x}-k^P)+(1-e^{n*})[1+\delta(1-F(\tilde{x}^w|\underline{\theta}))](b^P\underline{x}-k^P) \geq U^{Pw}$. But by the optimality of $\tilde{x}^n$, $U^{Pn} \geq U^{Pn'}$: contradiction. Thus $U^{Pw} \leq U^{Pn}$.

If M is conservative, then by Proposition 3 P's expected utility in Γ^w is:

$$U^P = \begin{cases} \delta e_\theta^{w*}(b^P\overline{x} - k^P) & \text{if } \overline{p} < \delta l(\underline{\theta}) \\ \delta[e_\theta^{w*}(b^P\overline{x} - k^P) + (1-e_\theta^{w*})(b^P\underline{x} - k^P)] & \text{if } \overline{p} \geq \delta l(\underline{\theta}) \text{ and } \frac{\delta(b^E\overline{x}-k^E)-\overline{p}}{2c} \geq \tilde{\mu} \\ 0 & \text{if } \overline{p} \geq \delta l(\underline{\theta}) \text{ and } \frac{\delta(b^E\overline{x}-k^E)-\overline{p}}{2c} < \tilde{\mu}, \end{cases} \tag{32}$$

where $e_\theta^{w*} = \max\left\{0, \frac{\delta(b^E\overline{x}-k^E)-\overline{p}}{2c}\right\}$. By Proposition 1 P's expected utility in Γ^n is $U^P = \delta[e^{n*}(b^P\overline{x}-k^P) + (1-e^{n*})(b^P\underline{x} - k^P)]$ if $\frac{\delta(b^E\overline{x}-k^E)-\overline{p}}{2c} \geq \tilde{\mu}$, and $U^P = 0$ otherwise, where $e^{n*} = e_\theta^{w*}$. The result then follows from the fact that $e_\theta^{w*}(b^P\overline{x} - k^P) > e_\theta^{w*}(b^P\overline{x} - k^P) + (1-e_\theta^{w*})(b^P\underline{x} - k^P) \geq 0$. ∎

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 $p^*(x_1', \theta', \emptyset) > \overline{p}^w$ 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 (conservative M) case, by assumption $x^M > \underline{x}$. It is then clear from Lemma 4 that M's report r ensures that $\mu < \tilde{\mu}$, 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 $x_1 = x_1'$.

Consider the alternate punishment strategy $p(x_1', \theta', \emptyset) = p'$. If E whistleblows under the alternate strategy, then M receives the same payoff as under $p^*(\cdot)$. If E does not whistleblow under the alternate strategy, then M receives a weakly higher payoff than under $p^*(\cdot)$, since M could have received the payoff from $p^*(\cdot)$ by reporting $r = \theta$. ∎

Proof of Comment 3. (i) With an aggressive manager, E's effort is $e^{w*} = \frac{b^E(\overline{x}-\underline{x})+\delta(b^E\overline{x}-k^E)+\overline{p}}{2c}$, which is clearly increasing in $\overline{p}$.

With a conservative manager, if $\overline{p} < \delta l(\underline{\theta})$, then $e^{w*} = \max\{0, \frac{\delta(b^E\overline{x}-k^E)-\overline{p}}{2c}\}$, which is clearly weakly decreasing in $\overline{p}$. If $\overline{p} \geq \delta l(\underline{\theta})$, then by Proposition 3 $e^{w*} = \begin{cases} \frac{\delta(b^E\overline{x}-k^E)-\overline{p}}{2c} & \text{if } \frac{\delta(b^E\overline{x}-k^E)-\overline{p}}{2c} \geq \tilde{\mu} \\ 0 & \text{otherwise.} \end{cases}$

Thus for $\overline{p}$ such that $\frac{\delta(b^E\overline{x}-k^E)-\overline{p}}{2c} \geq \tilde{\mu}$, e^{w*} is decreasing in $\overline{p}$ and strictly higher than any e^{w*} when $\overline{p} < \delta l(\underline{\theta})$. For larger values of $\overline{p}$ such that $\frac{\delta(b^E\overline{x}-k^E)-\overline{p}}{2c} < \tilde{\mu}$, Proposition 3 implies that $e_w^{w*} = 0$ for all $\overline{p}$. Thus e_w^{w*} is weakly decreasing in $\overline{p}$ for all $\overline{p}$. ∎

(ii) If M is aggressive and $\overline{p} < \delta l(\underline{\theta})$, then P's expected utility is: $U^P = e^{w*}(1+\delta)(b^P\overline{x} - k^P) + (1-e^{w*})(b^P\underline{x} - k^P)$. Differentiating yields $\frac{dU^P}{de^{w*}} = (1+\delta)(b^P\overline{x} - k^P) - (b^P\underline{x} - k^P)$. Since $b^P\underline{x} - k^P < 0 < b^P\overline{x} - k^P$, $\frac{dU^P}{de^{w*}} > 0$. By Comment 3, e^{w*} is weakly increasing in $\overline{p}$, thus U^P is weakly increasing in $\overline{p}$ for $\overline{p} < \delta l(\underline{\theta})$.

If M is aggressive and $\overline{p} \geq \delta l(\underline{\theta})$, then let $U^P(e)$ and $\tilde{x}^w(e)$ denote P's equilibrium expected utility and cutoff standard, respectively, when $e^{w*} = e$. P's expected utility is then $U^P(e) = e(1+\delta)(b^P\overline{x}-k^P)+(1-e)[1+\delta(1-F(\tilde{x}^w(e)|\underline{\theta}))](b^P\underline{x}-k^P)$. It is sufficient to show that for any $e' < e''$, $U^P(e') \leq U^P(e'')$. Suppose to the contrary that $U^P(e') > U^P(e'')$. Then by using standard $\tilde{x}^w(e')$ when $e^{w*} = e''$, P receives $U^{P'}(e'') = e''(1+\delta)(b^P\overline{x}-k^P)+(1-e'')[1+\delta(1-F(\tilde{x}^w(e')|\underline{\theta}))](b^P\underline{x}-k^P) > U^P(e')$. But by the optimality of $\tilde{x}^w$, $U^P(e'') \geq U^{P'}(e'')$: contradiction. Thus U^P is weakly increasing in $\overline{p}$ for $\overline{p} \geq \delta l(\underline{\theta})$.

To show that U^P is not increasing in $\overline{p}$ at $\overline{p} = \delta l(\underline{\theta})$, it is sufficient to show that $e^{w*}(1+\delta)(b^P\overline{x}-k^P)+(1-e^{w*})(b^P\underline{x}-k^P) > e^{w*}(1+\delta)(b^P\overline{x}-k^P)+(1-e^{w*})[1+\delta(1-F(\tilde{x}^w|\underline{\theta}))](b^P\underline{x}-k^P)$. Simplifying yields $0 > (1-e^{w*})\delta(1-F(\tilde{x}^w|\underline{\theta}))(b^P\underline{x}-k^P)$, which follows from the fact that $b^P\underline{x}-k^P < 0$.

If M is conservative, then P's expected utility is given by (32). Note that e_θ^{w*} is weakly decreasing in $\overline{p}$ and all three expressions for U^P in (32) are weakly increasing in e_θ^{w*}. Now for all $\overline{p} < \delta l(\underline{\theta})$, U^P is clearly weakly decreasing in $\overline{p}$. For $\overline{p} \geq \delta l(\underline{\theta})$, note that at $\overline{p} = \delta l(\underline{\theta})$, $e_\theta^{w*}(b^P\overline{x}-k^P) > e_\theta^{w*}(b^P\overline{x}-k^P) + (1-e_\theta^{w*})(b^P\underline{x}-k^P)$ and $e_\theta^{w*}(b^P\overline{x}-k^P) \geq 0$. Therefore, U^P is weakly decreasing in $\overline{p}$ over all $\overline{p}$. ■

Proof of Comment 4. It is straightforward (but cumbersome) to verify that for any π, the strategies characterized by (5), (7), Lemma 3, and Lemma 4 continue to hold, and that there exists a cutoff standard $\tilde{x}^\pi$ at which $\mu = \tilde{\mu}$ if $x_1 = \tilde{x}^\pi$.

If M is aggressive, then E's objective when M punishes only by type is modified from (14) as follows: $U^E = b^E(e\overline{x} + (1-e)\underline{x}) - k^E + \delta e(b^E\overline{x} - k^E) - (1-e)[\overline{p} - (1-F(\tilde{x}^\pi))\pi(b^M\underline{x} - k^M)] - ce^2$. When M punishes by whistleblowing and type, E's objective is modified from (16) as follows:

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

which evaluates to: $(1+\delta)[b^E(e\overline{x}+(1-e)\underline{x})-k^E]-\delta(1-e)(b^E\underline{x}-k^E)-(1-e)[\overline{p}-(1-F(\tilde{x}^\pi))\pi(b^M\underline{x}-k^M)] - ce^2$. Note that both objectives differ from (14) and (16) only through the addition of $(1-e)(1-F(\tilde{x}^\pi))\pi(b^M\underline{x} - k^M)$. Straightforward optimization yields the optimum interior effort level of $\frac{b^E(\overline{x}-\underline{x})+\delta(b^E\overline{x}-k^E)+\overline{p}-(1-F(\tilde{x}^\pi))\pi(b^M\underline{x}-k^M)}{2c}$. Thus effort is identical under either punishment strategy, and e^{w*} is weakly decreasing in π.

If M is conservative, then she may deter whistleblowing if $\overline{p} \geq \delta l(\overline{\theta}) + \pi|b^M\overline{x} - k^M|$. Following the derivation from Section 3.2, E's objective under this punishment strategy is identical to that in Γ^w, and thus her effort is the same as in (21): $e_w^{\pi*} = \begin{cases} \frac{\delta(b^E\overline{x}-k^E)-\overline{p}}{2c} & \text{if } \frac{\delta(b^E\overline{x}-k^E)-\overline{p}}{2c} \geq \tilde{\mu} \\ 0 & \text{otherwise.} \end{cases}$ If M punishes only by type, then her objective is: $U^E = -e(\overline{p} - (1 - F(\tilde{x}^\pi))\pi|b^M\overline{x} - k^M|) + \delta e(b^E\overline{x} - k^E) - ce^2$. (Note that $\tilde{x}^\pi$ may have a different value than in the aggressive manager case.) Straightforward optimization yields effort $e_\theta^{\pi*} = \max\left\{0, \frac{\delta(b^E\overline{x}-k^E)-\overline{p}+(1-F(\tilde{x}^\pi))\pi|b^M\overline{x}-k^M|)}{2c}\right\}$. By the argument in Proposition 3, M deters whistleblowing whenever feasible. Thus, the equilibrium effort level is $e_w^{\pi*}$ for π sufficiently low, and $e_\theta^{\pi*}$ otherwise. Combining expressions, it is clear that $e^{\pi*}$ is weakly increasing in π. ■

REFERENCES

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

Banks, Jeffrey S. 1989. "Agency Budgets, Cost Information, and Auditing." *American Journal of Political Science* 33(3): 670-699.

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): 809-839.

Bowman, James S. 1983. "Whistle Blowing: Literature and Resource Materials." *Public Administration Review* 43(3): 271-276.

Carpenter, Daniel P., and Michael M. Ting. 2006. "Regulatory Errors Under Two-Sided Uncertainty." Unpublished manuscript, Columbia University.

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): 224-244.

Gailmard, Sean. 2003. "Multiple Principals and Outside Information in Bureaucratic Policy Making." Unpublished manuscript, Northwestern University.

Gailmard, Sean, and John W. Patty. 2004. "Slackers and Zealots: Civil Service, Policy Discretion, and Bureaucratic Capacity." Unpublished manuscript, Northwestern University.

Gibbons, Robert. 1998. "Incentives in Organizations." *Journal of Economic Perspectives* 12(4): 115-132.

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): 421-435.

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.

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

Lewis, David. 2001. "Whistleblowing at Work: On What Principles Should Legislation Be Based?" *Industrial Law Journal* 30(2): 169-193.

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

Lewis, David E. 2004. "Presidents and the Politicization of the Institutional Presidency." Unpublished manuscript, Princeton University.

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): 243-277.

Melumad, Nahum D., Dilip Mookherjee, and Stefan Reichelstein. 1995. "Hierarchical Decentralization of Incentive Contracts." *Rand Journal of Economics* 26(4): 654-672.

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): 507-526.

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): 364-378.

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.

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

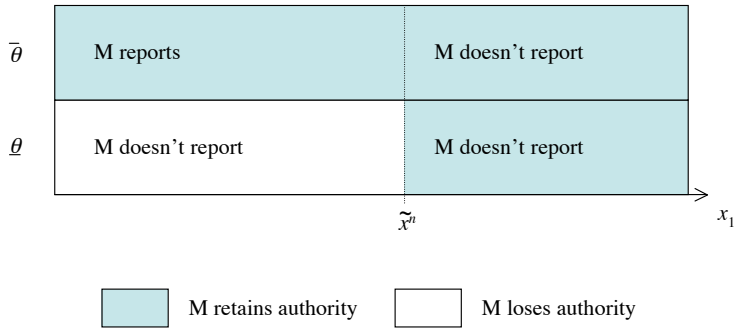

Figure 1: *No whistleblowing with an aggressive manager.* When M can report but E cannot whistleblow, a manager who wishes to approve all projects allows P to infer type $\overline{\theta}$ when x_1 is high. When x_1 is low, P infers $\underline{\theta}$ unless M reports. Thus P does not always learn θ when $\theta = \underline{\theta}$. P allows M to retain control unless x_1 is low and M is silent.

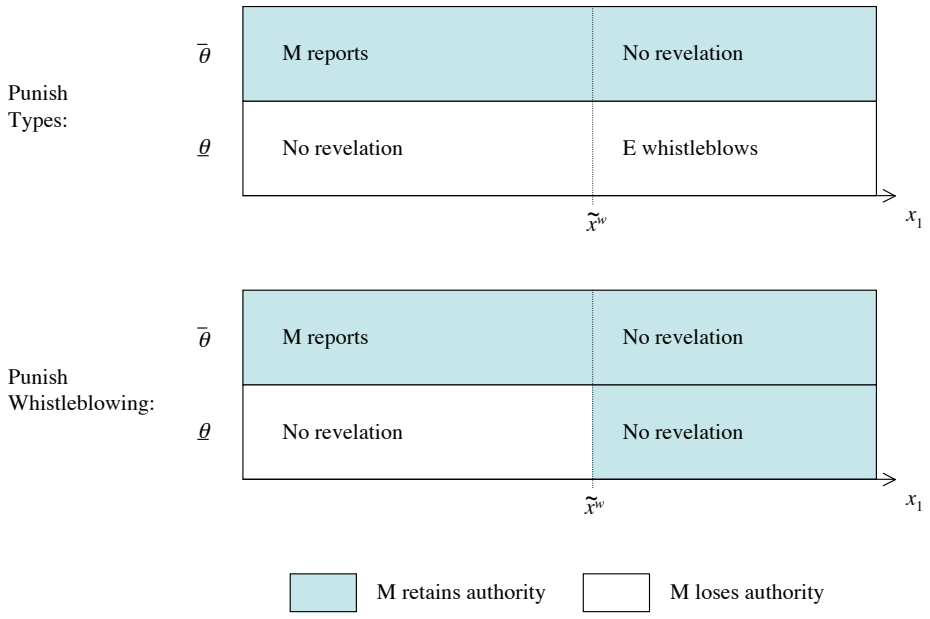

Figure 2: *Whistleblowing with an aggressive manager.* An aggressive manager will choose the same action as the politician when $\theta = \bar{\theta}$. If M punishes types, then E is free to whistleblow, and does so when x_1 is high and $\theta = \underline{\theta}$. If M punishes whistleblowing, then behavior resembles the no-whistleblowing case and P may be deceived about θ. Note that $\tilde{x}^w$, the outcome at which P infers $\theta = \underline{\theta}$, may be different from $\tilde{x}^n$.

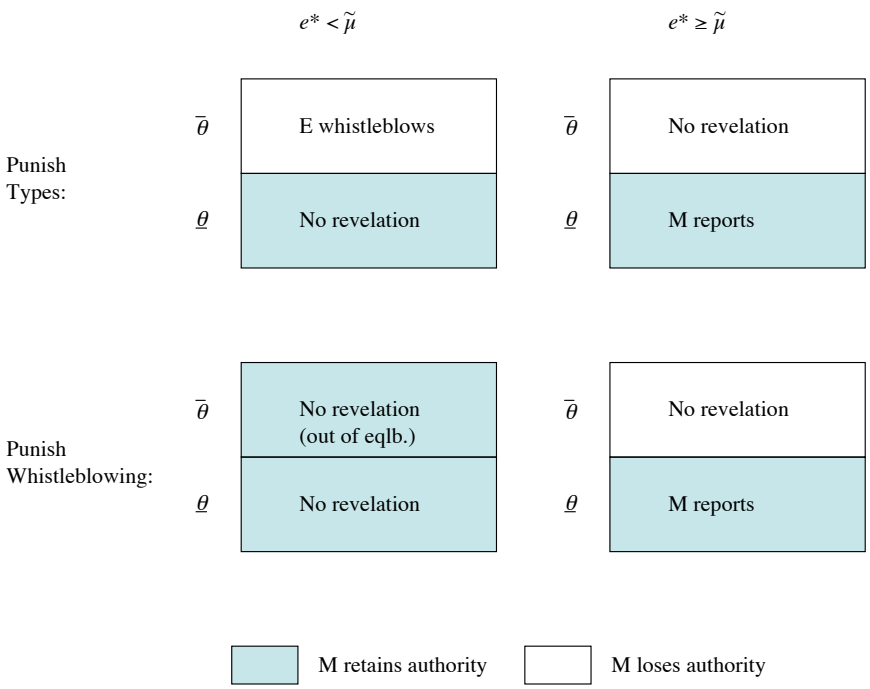

Figure 3: *Whistleblowing with a conservative manager.* A conservative manager will choose the same action as the politician when $\theta = \underline{\theta}$. M will cancel all period 1 projects, and therefore P must infer θ from effort e. In equilibrium, either E or M will have an incentive to reveal θ, and thus P always learns θ under both strategies.