Corporate Virtue: Motives for Whistle Blowing and the Proportionality of Punishment for Violators

Daniel G. Arce
School of Economic, Political and Policy Sciences
University of Texas at Dallas
2601 N. Floyd Road
Richardson, TX 75083-0688
USA
darce@utdallas.edu
Phone: 1-972-883-6725
Fax: 1-972-883-6297

Abstract

An evolutionary game-theoretic model is employed to address three essential aspects of whistle blowing: ethical decision making, the duality of mutual accountability among cohorts in large organizations, and role conflict between individual and organizational values. The extent of violations and proportionality of punishment are shown to depend upon two dimensions of role conflict. The first is the whistle blower's motivation: to provide a public good or avoid guilt by association. The second is the organizational culture regarding disclosure: hero or rat? Further, the model facilitates an evaluation of the whistle blowing provisions in the Sarbanes-Oxley (2002) Act.

Employees with insufficient virtue have done far more damage than those with insufficient talent.

–Charles Koch (2007), CEO of Koch Industries.

I. Introduction

Whistle blowing is an important phenomenon in public affairs and corporate America. In 2002, *Time Magazine* named whistle blowers Cynthia Cooper (WorldCom), Coleen Rowley (the FBI) and Sherron Watkins (Enron) its "Persons of the Year." Attitudes toward whistle blowing have been studied in-depth via survey methods in organizational behavior and management. By contrast, the economic literature on whistle blowing is sparse (Arce 2004); with the economics of crime focusing instead on hierarchical situations of law enforcement, including plea bargaining; or forms of information revelation by a party who is *a priori* guilty of a violation, such as cartel defection for leniency in sentencing. It is the purpose of this paper to examine ethical motives for whistle blowing within an evolutionary game-theoretic model of organizational culture.

Our working definition of whistle blowing stems from Bowie (1982) and Jubb (1999). It is a(n) (often public) voluntary action of organizational dissent, by an individual who has privileged information, in response to actual or suspected wrongdoing. The whistleblower has sufficient information to believe that disclosure constitutes convincing evidence to a reasonable person of the occurrence of wrongdoing. It is an indictment rather than merely informing. At the same time, whistle blowers generally do not have hierarchical control over the violator. Disclosure therefore involves *role conflict* between individual and organizational values. Individual motives for whistle blowing include maintaining personal integrity; avoiding complicity; and the need for action to remove the public 'bad' created by violators' actions.[1] Organizational role considerations include loyalty, obedience (complying with reasonable objectives), trust, and confidentiality. As Jubb (1999) emphasizes, whatever else it may be,

[1] An alternative investigation that differs from ours in that it focuses on whistle blowing as a preemptive action in response to *anticipated* wrongdoing is given by Heyes (2005).

whistle blowing presents an ethical dilemma originating in role conflict. That is, is the whistle blower a hero or rat?

When the attitudes towards whistle blowers (hero or rat) are reconciled with the whistle blower's rationale for disclosure (creating a public good versus avoiding guilt by association) the result is an equilibrium characterization of the proportionality of punishment. Our analysis therefore combines literatures stemming at least from Brams and Kilgour (1985, 1987) and Casson (1991) on the proportionality of punishment and the evolution of corporate culture. In particular, we examine the outcomes of social learning within large corporate/bureaucratic entities as if they are outcomes of an evolutionary process. Further, within large organizations, such as those involving Cooper, Rowley, and Watkins, a whistle blower is unlikely to have hierarchical authority over ethical violators. Hence, in contrast to principal-agent theory, where the principal's claim over productive surplus motivates his/her monitoring function, whistle blowers' decisions are based on an examination of the tradeoff between their personal assessment of ethical misbehavior, and the degree to which organizational culture indicates how their disclosure will be treated (Gundlach et al 2003). Moreover, whereas government whistle blowers have protection from reprisals through whistle blower protection acts, prior to the Sarbanes−Oxley Act (2002) the same could not have been said of whistle blowers in public corporations. Consequently, a need to examine provisions in Sarbanes−Oxley (hereafter SOX) exists so as to ascertain the effect of SOX on organizational culture and whistle blowing.

The paper proceeds as follows. In the next section we introduce an enforcement game without hierarchy. The degree of misbehavior and the organizational culture regarding whistle blowers determine the incidence of reporting *and* the information structure itself, which is defined by the amount of cohort monitoring in equilibrium. Further, the lack of hierarchy implies a degree of duality not present in agency models. This duality recognizes that whistle blowing often takes

place within a context of a role-contingent strategy choice where cohorts decide whether to (i) monitor or not, and (ii) violate or adhere to an ethical norm. Indeed, this duality is what facilitates a single-population evolutionary analysis of the enforcement game (von Wangenheim 2004). It also restores the link between punishment and deterrence that is missing in enforcement games, but is present in Becker's (1968) non-strategic analysis, thereby reconciling the two approaches to crime and punishment. Section III formally addresses the issue of corporate culture as an evolutionary outcome. As a result, whistle blowing is not required to be an absolute ethic. Instead, we derive an equilibrium distribution of ethical and unethical agents that is a function of the two critical dimensions outlined above: organizational culture (hero or rat) and the whistle blower's personal motivation for disclosure. In section IV we characterize the proportionality of punishment for violators as a function of these two critical dimensions. Specifically, we derive the conditions for disproportionate punishment versus *lex talionis* (an-eye-for-an-eye) and also evaluate the degree to which of SOX addresses false accusations. The final section contains brief concluding remarks.

II. Whistle Blowing: Monitoring Without Hierarchy

In this section we present an evolutionary game theoretic model that that is constructed to address three essential features of whistle blowing: ethical decision making, the duality of mutual accountability among cohorts in large organizations, and role conflict between individual and organizational values. The analysis of organizational culture requires an emphasis on behavior within a large group of players (Casson 1991). In each period a member of the group experiences one encounter with a randomly selected cohort. Box 1 specifies the outcome of any such pairwise matching. Formally, strategies refer to player *types*; they are the units of selection that characterize the organization's behavior. From an evolutionary perspective, the real contest is a pairwise competition between different *strategies*, rather than individual players; i.e., Box 1

is *not* interpreted as a strategic form game. Further, *roles* refer to the row (monitor) or column (worker) position. In the monitor role, an employee can either monitor (M) or not (N). In the worker role an employee can violate (V) an ethical code or adhere to it (A). The fact that the whistle blower is not assumed to be a violator makes it explicit that we are not associating whistle blowing with pejorative stereotypes such as plea bargaining, cartel defection for leniency in sentencing, and other forms of information revelation by a party who is *a priori* guilty of a violation. Finally, payoffs are not expressed *ex ante* in terms of the population share of whistle blowers or violators because we endogenously derive these equilibrium proportions. *Expected* payoffs will therefore be a function of these population shares.

The payoffs in Box 1 are expressed in terms of the perceptions and consequences of ethical breaches. In the baseline case, represented by the southeast cell corresponding to a (N, A) pairing, no wrongdoing occurs, nor is there any determination of hero or rat. In this case the marginal payoff to each type of behavior is 0. In the (N, V) cell an ethical breach occurs. A violator commits wrongdoing in order to gain some kind of advantage, which we label as an ethical breach of value e. From the whistle blower's perspective this has a negative impact equal to $-x < 0$. For now, x is treated as the unknown rationale for whistle blowing, expressed as avoiding a payoff of $-x$. The two interpretations we investigate correspond to avoiding guilt by association versus minimizing the negative externality created by an ethical breach.

In an (M, V) matching wrongdoing is observed, as is standard in the literature on inspection games (Andreozzi 2004) and agency with monitoring (Arce 2007). This also follows because whistle blowing is justified from an ethical and legal perspective only if the whistleblower has made certain that he/she has evidence that is externally verifiable and convincing to a reasonable person (Bowie 1982, Gundlach et al 2003). Note that the equilibrium nature of the informational structure of monitoring is derived below. A whistle blower expects

the violator to incur a penalty, $P > 0$, given by the second entry in the northwest cell. The whistle blower receives payoff $H–m$, where H is potentially a 'hero's' payoff, but could also involve retaliation, and $m > 0$ is the cost of monitoring. *A priori* we require:

(1) $H–m > -x$;

i.e., it is preferable to blow the whistle if the violation is sizeable enough. If the damage due to wrongdoing does not exceed $H–m$, no whistle is blown. In a standard agency relationship, H represents the principal's hierarchical claim on the residual surplus. The principal would then set P to maximize its residual surplus. In our situation no such hierarchy or claim exists; instead, the relationship between H and P is endogenously determined by an equilibrium process that explicitly recognizes the absence of hierarchy. By contrast, in enforcement games no direct relationship between H and P exists (Andreozzi 2004).

Another possibility is an (M, A) pairing, where the row player is unnecessarily monitoring, thereby incurring a (possibly social) cost among his/her cohorts. Such behavior typifies a 'rat,' with the whistle blower incurring $–R < 0$ in addition to the monitoring costs. As no violation has occurred, the column player faces a false accusation of value $–F < 0$. One objection to SOX is that it may lead to type II errors in the form of harmful false accusations. We assume that:

(2) $0 \leq F < P$.

This inequality states that, when considering adherence versus violating, the penalty for a violation is a greater deterrent than the stigma association with a false accusation. Indeed, the latter is a form of type II error, the equilibrium occurrence of which is characterized below.

The payoff structure described above implies that the Nash equilibrium for Box 1 occurs in mixed strategies, as is well-known for inspection games. Given that the evolutionary equilibrium concepts we employ are refinements of Nash equilibrium, they posses this property as well. As in

Selten (1980), all matchings are assumed to be nonassortative – there are no matchings among players in the same role (monitor-monitor or worker-worker) − because our focus is on the whistle blowing ethic as defined by a worker-monitor matching among cohorts. Consequently, Box 1 is also a subgame of Figure 1, which specifies the random process of nonassortative matching. With probability .5 nature, $\mathcal{N}$, selects an employee into the monitor role and matches him/her with a cohort in the worker role. Correspondingly, with probability .5 the reverse role assignment is made. This model is therefore meant to address the Yin-Yang of mutual accountability that characterizes ethical systems.[2] Potential role reversals of this type require cohorts to internalize both the consequences of committing *and* reporting violations, a process that is absent in agency and inspection models.

Given that each subgame in Figure 1 is strategically identical to Box 1, it suffices to identify the Nash equilibrium for Box 1 in order to characterize the subgame perfect equilibrium (hereafter, SPNE) for Figure 1. We denote local/behavioral strategy λ_k as the frequency of behavior k, $k \in \{M, N, V, A\}$; where $\lambda_M + \lambda_N = 1$, $\lambda_V + \lambda_A = 1$. As detailed above, the corresponding equilibrium is in mixed strategies.

RESULT 1: the subgame perfect Nash equilibrium for Figure 1 is (proof in appendix):

(3) $\lambda_M = e/(P+e-F)$, $\lambda_N = (P-F)/(P+e-F)$; $\lambda_V = (R+m)/(H+R+x)$, $\lambda_A = (H+x-m)/(H+R+x)$.

Strategic analyses of deterrence and inspection games have been characterized by the counterintuitive result that the occurrence of monitoring/enforcement is unaffected by the cost or rewards for enforcement, H and m; and an increase in the severity of penalties, P, leaves the frequency of violations unchanged (Andreozzi 2004). By contrast, Becker's (1968) classic (non-

[2] Similarly, *Time Magazine* (2002) likens the whistle blower context to the statement on unwelcome truth telling in Ibsen's play, *An Enemy of the State*, "A community is like a ship (…) everyone ought to be prepared to take the helm."

game-theoretic) analysis recommends that P should be set as its maximum level because it is a more cost-effective method of deterrence than increasing the frequency of monitoring. The independence arises in a game-theoretic analysis because mixed strategies for one player are derived by solving a series of linear equalities for the expected payoffs of the *other* player. A similar independence holds for the retaliation functions in Brams and Kilgour's (1987) analysis of deterrence. Currently, our model as well exhibits this property: the proportion of monitors, $\lambda_M = e/(P+e-F)$, is not a function of H, R or m. Similarly, the proportion of violators, $\lambda_V = (R+m)/(H + R + x)$, is not a function of P. These missing connections are established below for an (evolutionary) environment that explicitly recognizes the Yin-Yang of role duality in organizational culture.

As was foreshadowed, the structure of information is endogenous:

COROLLARY: information about ethical violations is imperfect in that cohorts in the worker role are monitored with probability λ_M and not at all with probability λ_N $(= 1 - \lambda_M)$.

This corollary clarifies the nature of imperfect monitoring in our model. In agency models the agent's action is *assumed* to be hidden, and this information asymmetry drives the need for the principal's ability to (i) claim the residual profit derived from the agent's effort, and (ii) terminate the agent's contract, in order to align incentives within the firm. Technically, the agent's action is not unobservable, but in equilibrium it is too costly to monitor to the point of perfect information about the agent's action. By contrast, a cohort in the monitor role has *neither* of these abilities. Instead, our focus is on information asymmetry in the sense of Selten (1980), due to its correspondence with Jubb's (1999) argument that whistle blowing is all about role conflict, rather than imperfect information about agents' actions. Together, role conflict and the absence of hierarchy result in an *endogenous* relation between H, P, R, F, e and x.

III. Evolution and Organizational Culture

The effects of organizational culture on whistle blowing are an ongoing, rather than one-shot process. Evolutionary equilibrium concepts have been shown to formally correspond to the stable equilibria of replicator dynamics for strategies such as ours that are meant to characterize the behavior of a population of players (within a large organization). By replicator dynamic we mean the dynamic process in which the growth rate of a strategy within a population is equal to the difference between its expected payoff and the average expected payoff, weighted by the frequency of the strategy within the population.[3] Further, the conceptual link between game theory and evolutionary theory is the parallel between incremental learning and natural selection (Rapoport 1994). Humans are biologically equipped to learn. In this way, we are able to interpret our evolutionary environment in the context of cohorts learning/adapting to the relative merits of behavior, as determined by organizational culture (Casson 1991, Arce 2007, Kolstad 2007). Consequently, the results in this section do not correspond to a 2-player game, but rather a population characteristic through pairwise matching and the replicator dynamic.

Figure 1 corresponds to Selten's (1980) evolutionary analysis of asymmetric games. By asymmetric game Selten meant a situation where (i) the row (r) and column (c) strategy sets are not equal, $\Sigma_r \neq \Sigma_c$, and/or (ii) the player's (expected) payoffs, $E_r[\lambda_r, \lambda_c]$ and $E_c[\lambda_r, \lambda_c]$, are not interchangeable if row and column exchange strategies: $E_r[\lambda, \lambda'] \neq E_c[\lambda', \lambda]$ for some strategy pair (λ, λ'). In a symmetric game the "$\neq$" in both (i) and (ii) is replaced with "=" for all strategy pairs. Evolutionary game theory is based on the notion of symmetry because it is meant to identify the traits that emerge from a homogeneous population of players through the process of natural selection via the replicator dynamic (Hirshleifer 1977). The transformation given in Figure 1 was

[3] Formally, if σ is the current joint mixed strategy/distribution for the game and σ_s corresponds to the frequency of pure strategy s, then the growth rate of s is given as $\dot{\sigma}_s = \sigma_s\{E[s, \sigma] - E[\sigma, \sigma]\}$.

created by Selten to symmetrize asymmetric games. As role assignment is uncertain *a priori*, players select *role-contingent* strategies: $\lambda_r\lambda_c \in \Delta(\Sigma_r)\times\Delta(\Sigma_c)$. Expected payoffs are expressed as:

(4) $\quad E[\lambda_r\lambda_c, \hat{\lambda}_r\hat{\lambda}_c] = .5E_r[\lambda_r, \hat{\lambda}_c] + .5E_c[\hat{\lambda}_r, \lambda_c].$

The coefficients of .5 in (4) can be dropped without loss of generality. When applied to Box 1 the resulting game is illustrated in Box 2. For example, in a pairwise matching between a *MA* type and an *NV* type, the payoffs in (4) are taken from Box 1. The expected payoff for the *MA* type is $E[MA, NV] = E_r[M, V] + E_c[N, A] = H–m + 0 = H – m$, corresponding to the first entry in the (*MA*, *NV*) cell in Box 2. This game – known as an *asymmetric contest* – represents the transformation of Box 1 via Figure 1. It is symmetric; hence, evolutionary concepts of equilibrium can be applied, thereby facilitating an evolutionary analysis of whistle blowing, ethics, and organizational culture.

The equilibrium concept we employ is the neutrally stable strategy (hereafter, NSS), as defined by Maynard Smith (1982: 107).

DEFINITION: strategy σ is a *neutrally stable strategy* (NSS) for a symmetric game if:

(a) $\quad E[\sigma, \sigma] \geq E[\sigma', \sigma] \; \forall \; \sigma' \in \Delta(\Sigma)$, where $\Sigma = \Sigma_r = \Sigma_c$.

(b) $\quad$ If $E[\sigma, \sigma] = E[\sigma', \sigma]$ then $E[\sigma, \sigma'] \geq E[\sigma', \sigma']$.

Condition (a) requires σ to be a symmetric Nash equilibrium. Condition (b) is an additional stability requirement specifying that if σ' is an alternative best reply to σ, then σ must be at least as good a reply to σ' as σ' is to itself.[4]

[4] In comparison, an *evolutionary stable strategy* (ESS) requires (b) to hold with strict inequality (> rather than ≥) if (a) holds with equality. Selten's (1980) theorem establishes that evolutionary stable strategies for asymmetric games are never mixed. Hofbauer and Sigmund (1998: 125) provide the theorem that σ is an ESS for the asymmetric contest iff it is strict pure strategy Nash equilibrium (involving unique mutual best replies) for the original asymmetric game. Our game has no ESS; Box 1 has no strict Nash equilibria under the parameter values investigated. There is no general equilibrium result for NSS's either.

Neutral stability is akin to Noreen's (1988) requirement that corporate ethics need not be absolute in that everyone adheres to them, but once a critical mass adopts them, sufficient social reinforcement exists for members to benefit by adhering to them. Neutral stability been used to demonstrate how norms of cooperation (Fudenberg and Maskin 1990) and fairness (Ellingsen 1997) can arise as outcomes of an evolutionary process. Moreover, an NSS in mixed strategies can be interpreted as a *polymorphism*: a cross-sectional distribution of player types among cohorts. In this way, neutral stability does not require uniform behavior across all members; a small subgroup can exist within the organization that achieves similar ends (expected payoffs) via different means. Kolstad (2007) gives the example that an informal mode of organization may not resist invasion by a formal one, but might nevertheless be stable if the formal mode does not yield strictly higher expected payoffs. From a cultural perspective, the best an incumbent norm can be expected to do is limit the further spread of payoff-equivalent alternatives, rather than rule them out entirely. This corresponds to the property that an NSS is Lyapunov stable with respect to the replicator dynamic (specified in footnote 3) when applied to any pairwise contest in Box 2, meaning that any perturbation from the NSS will not move any further from the NSS (and may return to it). For evolutionary stability, no alternative strategy can persist; for neutral stability, no alternative strategy can thrive (increase their population share).

In conclusion, the environment/equilibrium concept we employ has the following advantages: (i) it is a static characterization of an evolutionary outcome consistent with a learning process that culminates in an organizational culture; (ii) agents are self-reflecting in that they consider the consequences of their actions in each role (monitor or worker) due to the absence of hierarchy; (iii) it similarly captures the role conflict inherent in whistle blowing (hero versus rat); and (iv) the stability of a norm does not necessary imply uniform behavior, in that polymorphisms are allowed and norms are robust against payoff-equivalent deviations. As a

result, we are able to establish equilibrium relations between the consequences for violating and reporting violations, and endogenously derive the proportionality of punishment.

IV. Equilibrium Implications of Role Duality

In this section we derive an evolutionary equilibrium characterization of the whistle blowing scenario in Box 1. The presence of role duality has two important implications. First, it facilitates a non- hierarchical analysis, consistent with evolutionary models of a single population (von Wangenheim 2004). Second, players must consider the consequence of their actions in adhering to and enforcing the social norm, thereby breaking the independence between the penalty for a violation and its frequency that so often characterizes strategic enforcement/deterrence theory. Role duality requires each player to *internalize* the consequences of behavior in *both roles*. For example, in the asymmetric contest (Box 2) both the reward for monitoring and the punishment for a violation are present in the payoffs in the (*MV*, *MV*) cell. The resulting equilibrium produces an endogenous relation between the consequences for monitoring and those for violations, and gives prescriptions for the proportionality of punishment.

In an asymmetric contest σ_{MV} is the population proportion of *M*-types in the monitor role and *V*-types in the worker role, σ_{MA} is the proportion of joint *M*- and *A*-types, etc. When NSS is applied to the asymmetric contest we identify an outcome where organizational members consider their dual role as monitor and worker (via role-contingent strategies), learn from the organizational culture the relative merits of different types of behavior, and is characterized by the population proportion that adheres to them as a function of past behavior (through the replicator dynamic). Further, because the transformation used to create the asymmetric contest in Box 2 − Figure 1 − is of perfect recall (no *ex post* uncertainty over role assignment exists), it holds that $\sigma_{ij} = \lambda_i\lambda_j$ ($i = M, N; j = V, A$) where λ_i and λ_j are the local strategies in Result 1.

A well-known property of mixed strategy Nash equilibrium is that any pure strategy, s, that is played with positive probability, $\sigma_s > 0$, has the same expected payoff as any other pure strategy played with positive probability, and the mixture itself (e.g., $E[s, \sigma] = E[\sigma, \sigma]$). Given the correspondence between the mixed strategy SPNE for Figure 1 and the Nash equilibrium for Box 2 discussed above, $\sigma_{ij} = \lambda_i\lambda_j$, it follows that NSS condition (a) holds with equality for (role-contingent) pure strategies $s \in \{MV, MA, NV, NA\}$ because each occurs with positive frequency in the equilibrium given in (3). NSS condition (b) must therefore be examined to characterize the equilibrium. For example, it must be the case that $E[\sigma, MV] \geq E[MV, MV]$; i.e.:

$$[H–P–m]\sigma_{MV} + [H–F–m]\sigma_{MA} – [P+x]\sigma_{NV} – [F+x]\sigma_{NA} \geq H – P – m.$$

Aggregating terms:

$$[H–m](\sigma_{MV} +\sigma_{MA}) – P(\sigma_{MV} +\sigma_{NV}) – F(\sigma_{MA} +\sigma_{NA}) – x(\sigma_{NV} +\sigma_{NA}) \geq H – P – m.$$

Under perfect recall, $\sigma_{ij} = \lambda_i\lambda_j$. Further, $\lambda_M + \lambda_N = 1$ and $\lambda_V + \lambda_A = 1$, imply:

$$[H–m]\lambda_M – P\lambda_V – F\lambda_M – x\lambda_N \geq H – P – m.$$

Once again aggregating terms: $[P–F]\lambda_A \geq [H+x–m]\lambda_N$. Substituting in the values of λ_A and λ_N from (3) yields $P + e – F \geq H + R + x$.

Continuing, it must also be the case that $E[\sigma, MA] \geq E[MA, MA]$. That is:

$$–[R+P+m]\sigma_{MV} – [R+F+m]\sigma_{MA} – P\sigma_{NV} – F\sigma_{NA} \geq –(R + F + m).$$

Aggregating terms: $–[R+m](\sigma_{MV} +\sigma_{MA}) – P(\sigma_{MV} +\sigma_{NV}) – F(\sigma_{MA} +\sigma_{NA}) \geq –(R + F + m)$. Perfect recall, $\sigma_{ij} = \lambda_i\lambda_j$, and the add-up conditions on the λ_i's yields $[R+m]\lambda_N \geq [P–F]\lambda_V$. Under (3) this becomes $H + R + x \geq P + e – F$. By similar method $E[\sigma, NV] \geq E[NV, NV]$ yields $H + R + x \geq P + e – F$ and $E[\sigma, NA] \geq E[NA, NA]$ yields $P + e – F \geq H + R + x$. Together, these inequalities imply $H + R + x = P + e – F$.

RESULT 2: The (unique) equilibrium in (3) is an NSS for Box 2 if:

(6) $H + R + x = P + e - F$.

Equation (6) aggregates our measures of organizational culture into one equilibrium condition. Further, this characterization is a consequence of the evolutionary paradigm, owing to its specific recognition of the role conflict embodied in whistle blowing. In particular, role-contingent strategy choice requires organizational members to internalize the consequences of actions in *both* roles. Hence, by explicitly recognizing the Yin-Yang of role conflict in an ethical system we have established a strict relationship between the consequences for the whistle blower – H, R and x – and those for his/her cohort – P, F and e – previously absent in inspection games. This has implications for the proportionality of punishment, as investigated below.

V. Motives for Whistle Blowing and Attitudes toward Disclosure

We examine whistle blowing behavior in two dimensions. The first is the whistle blower's ethical rationale for disclosure. This is captured by the $-x$ payoff that the whistle blower seeks to avoid in a (N, V) pairing. The second is the organizational culture regarding disclosure; is the whistle blower viewed as a hero or rat? A novel result is that when these dimensions are considered simultaneously, we can characterize the proportionality of punishment. Further, as Sarbanes-Oxley (SOX) addresses the treatment of whistle blowers, we can evaluate the effectiveness of this piece of legislation within our characterization; particularly the potential for false accusations.

Consider first our potential rationales for whistle blowing. Whereas a violator is motivated by personal gain, e, some whistle blowers recognize that the damages stemming from the violation affect a larger constituency (Brewer and Selden 1998). In particular, *Time*

Magazine's 2002 Persons of the Year blew the whistle on actions related to terror interdiction and corporate malfeasance. Terrorism creates a public bad because the violence associated with an attack is meant to intimidate an audience beyond that of the immediate victims. Similarly, a broad range of stakeholders are affected by corporate malfeasance. In terms of our model, the public aspects of whistle blowing can be captured by setting $x = ne$, where $n \geq 2$ is the size of the constituency affected by the ethical breach. The breach has value e for the violator, and creates a negative externality equal to $-ne$. Whistle blowers with a public good motivation therefore seek to negate the creation of negative externality ne.

Jubb (1999) describes an alternative moral perspective for the whistle blower, one in which tolerance of the violation is tantamount to compliance and warrants similar punishment if discovered (guilt by association). In this way, $x = P$. Here, whistle blowers are motivated by an ethic in which failure to monitor will leave them equally complicit, and thereby susceptible to the same punishment received by the violator, P. In this approach, whistle blowers aim to avoid complicity and maintain personal integrity.

The second dimension of whistle blowing is the corporate culture regarding disclosure. We consider three possibilities. The first is that the whistle blower is a rat; $H = -R$ in terms of our model. Setting $H = -R$ allows us to examine the consequences for organizations that do not recognize the importance of virtue in their success, as contrasted with the quote by Charles Koch (2007) that opens this paper. In other words, the organization fails to honor moral commitment. Our second case is that of a *type II* organization, referring to the statistical concept of avoiding false positives, which corresponds to the social effect of false accusations, R and F, in our model. Finally, we address the treatment of whistle blowers mandated in SOX. Several sections of SOX establish guidelines for civil awards to protect whistle blowers, and criminal felony penalties for retaliators. By law, SOX-regulated firms must have a no retaliation policy within their employee

manuals. In terms of our model, SOX is geared toward the ideal of $R = 0$. This interpretation is appropriate because SOX does not require the whistle blower's allegations to be correct, but rather that he/she had "reasonable cause" to believe that unlawful activity occurred.

The implications of combining the whistle blower's motivation with the organization's treatment of disclosure are summarized in table 1. The entries in this table are derived by applying the corresponding row and column headings to equilibrium condition (6): $H + R + x = P + e - F$.

For example, when a whistle blower is a rat, $H = -R$, and whistle blowers seek to avoid guilt by association, $x = P$, then applying these two equalities to $H + R + x = P + e - F$ yields $e = F$. As $P > F$, the corresponding entry, $P > e$, is derived for this cell. Violations receive a greater than proportional punishment. Indeed, Bowie (1982: 138) argues that failure to honor moral commitment within an organization, $H = -R$ here, undermines the organization's structure. The way to correct for this is to propose greater than proportional penalties, which serves as a successful deterrent. To see this, recall from result 1 that $\lambda_M = e/(P+e-F)$. Further, the proportion of violators is not a function of the penalty for a violation: $\lambda_V = (R+m)/(H+R+x)$. The equilibrium condition in (6) allows us to tie λ_V to P. In particular, when $H = -R$, $x = P$, and $e = F$; then $\lambda_M = e/P$ and $\lambda_V = (R+m)/P$. As the punishment, P, increases, λ_M and λ_V decrease, coming arbitrarily close to the ideal of (N, A). In this way, we have restored Becker's (1968) intuition about punishment and deterrence to the strategic analysis of enforcement games.

Consider the top entry in the second column of table 1. When $H = -R$ but now the whistle blower has a public good motive, $x = ne$, substitution of these equalities into equilibrium condition $H + R + x = P + e - F$ yields $P = (n-1)e + F$. Given that the negative externality associated with the violation is ne, the recommended punishment is greater than, equal to, or less than proportional depending upon whether $F > e$, $F = e$, or $F < e$, respectively. In other words,

the proportionality of punishment depends upon the consequences of a false accusation, F, in comparison with the incentive for an ethical breach, e. When the consequences of a false accusation are particularly onerous, $F > e$, it is as if there is a greater incentive to violate, rather than adhere. As a consequence, punishment will have to be greater than proportional in order to dissuade those in the agent role from committing a violation. Once again it is useful to investigate the limit properties of this equilibrium, which are now expressed in terms of the magnitude of the externality, n. Together, $H = -R$, $x = ne$, and $P = (n-1)e + F$ imply $\lambda_M = e/(P+e-F) = 1/n$ and $\lambda_V = (R+m)/(H+R+x) = (R+m)/ne$. As the size of the externality, n, increases, λ_M, and λ_V decrease, again coming arbitrarily close to the ideal of (N, A). This is because the requisite punishment, $P = (n-1)e + F$, increases with n.

When the whistle blower is viewed as a rat, punishment must be greater than proportional in order to deter violations because the disincentive to disclose, $H = -R$, means a lower incidence of monitoring. Hence, as in Brams and Kilgour (1985, 1987) disproportionate punishment is warranted. We have identified that when an organization lacks moral commitment, it will have to rely on greater-than-proportional punishment to deter violations.

RESULT 3: When a whistle blower is a rat, guilt by association implies a greater than proportional punishment. By contrast, if the motivation for whistle blowing is to provide a public good, then proportionality depends upon the relationship between the consequences of a false accusation, F, versus the incentive to violate, e. If $F > e$, then punishment should also be greater than proportional. In these cases increasing the punishment approaches the ideal of (N, V).

Brams and Kilgour (1985, 1987) further argue that proportional punishment itself corresponds to a special case. In our model these special cases can be derived by subtracting ($R+F$) from the incentive to disclose (monitor violations). Such an organizational culture is labeled as *Type II*; subtracting ($R+F$) demonstrates a concern for false positives, the social costs in the (M, A) cell of Box 1. False accusations are of concern because $\lambda_M \times \lambda_A > 0$ in equilibrium.

When $x = P$ we investigate a punishment-based point of reference, $H = P - (R+F)$. Substituting $H = P - (R+F)$ and $x = P$ into equilibrium condition $H + R + x = P + e - F$ yields $P = e$. Alternatively, when $x = ne$ our point of reference is violation-based, $H = e - (R+F)$. The 'e' term in $H = e - (R+F)$ corresponds to the whistle blower's "share" of the negative externality produced by a violation. Substituting $H = e - (R+F)$ and $x = ne$ into (6) derives $P = ne$. In summary, when the consequences of a violation are $x = e$ for the whistle blower, the appropriate punishment is $P = e$. When these consequences are $x = ne$, it is $P = ne$. For both cases of type II organization, *the punishment must fit the crime*.[5]

We have therefore derived the special cases corresponding to the well-known ethical maxim of 'an-eye-for-an-eye;' a form of retributive justice known as *lex talionis*, or law of equivalency. It is the punitive form of the golden rule. Having roots traced back at least to Hammurabi's code (Babylonia, circa 1870 BC), *lex talionis* is one of the world's first recorded laws. It is also found in Biblical, Islamic and Roman Law.[6] As in our model, *lex talionis* was originally intended for situations lacking hierarchy, such as the settlement of disputes between two families. It is not meant to be taken literally, but is instead an egalitarian standard intended to limit retaliation. The eye-for-an-eye principle places rational limits on vigilantism and revenge. It has a complex deterrence value; affirming that crime should not pay, but also forbidding excessive reprisals − particularly in the form of sending a message through symbolic punishment. Our model extends *lex talionis* to the case where the damages caused by the violation go beyond the immediate victim. When the motivation for whistle blowing is to avoid the negative externality, $x = ne$, proportionality corresponds to $P = ne$. By analogy, at the turn of the millennium many of those found guilty of crimes associated with corporate malfeasance were given sentences that

[5] For this reason we do not explore the functional relationship between λ_M, λ_V and P. The magnitude of P is fixed by proportionality.
[6] See, for example, Exodus 21:24 and the Qur′an 5:45.

exceeded the average sentence for murder in the corresponding jurisdiction. Such sentencing recognizes the extent to which these violations affected multiple stakeholders.

Finally, these conditions are novel in that they specifically *prescribe* tit-for-tat in equilibrium, rather than tit-for-tat being one in a plethora of potential solutions (e.g., via the folk theorem). The underlying logic is explained by the process of internalization in the asymmetric contest (Box 2). The Yin-Yang of role duality implies that members must consider the consequences of role reversal; i.e., reporting *and* violating. In particular, the (MV, MV) payoff in Box 2 is $H - P - m$; the punishment for a violation enters into a member's payoff when he/she is monitoring. The $-F$ term in the (MA, MA) cell is cause for self-reflection as well. Hence, in order to monitor/report a violation, the member must be able to internalize the consequences, thereby requiring that it be proportional to the violation.

RESULT 4: For type II organizations – those concerned with false accusations – specific whistle blower compensations exist that correspond to proportional punishment: $H = P - (R+F)$ for avoiding guilt by association, and $H = e - (R+P)$ for a public goods motive. In both cases this compensation is a negative function of the total consequences of false accusations, $(R+F)$.

We conclude with an examination of the whistle blowing provisions in SOX. Several sections of SOX establish guidelines for civil awards to protect whistle blowers, and criminal felony penalties for retaliators. By law, SOX-regulated firms must have a no retaliation policy within their employee manuals. In terms of our model, SOX is geared toward the ideal of $R = 0$. This interpretation is appropriate because SOX does not require the whistle blower's allegations to be correct, but rather that he/she had "reasonable cause" to believe that unlawful activity occurred. When $x = P$ and $R = 0$ are substituted into equilibrium condition $H + R + x = P + e - F$ the punishment for a violation is left unspecified. The punishment is no longer derived from equilibrium conditions; instead, it must be mandated from outside. As a guideline, simple

substitution of the values for this case yields $\lambda_M = e/(P+e-F) = e/(P+H)$ and $\lambda_V = (R+m)/(H+R+x) = m/(P+H)$. Increased punishment again reduces λ_M and λ_V, approaching the ideal of (N, A).

In addition, it is best to treat the whistle blower as a hero, $H > 0$, as λ_M and λ_V are lower when H is positive. Moreover, the whistle blower's reward is derived as $H = e - F$. The hero's reward is given by the difference between the violator's personal gain for the ethical violation, e, and the consequences of making a false accusation, F. Interestingly enough, this reward is equivalent to that derived to avoid type II errors, $H = e - F - R$, because $R = 0$. When $H = e - F$ it is possible for SOX to be consistent with *either* whistle blowing ethic given in the columns of table 1. Further, this leads to an intuitive prescription for punishing violators – proportionality.

SOX was created by Congress to quickly reassure constituents who were reluctant to reenter securities markets, owing to constituents' massive loss of wealth in the wake of blockbuster corporate scandals. Hence, from Congress' perspective whistle blowing provides a public good. Correspondingly, when $x = ne$ and $R = 0$ the equilibrium condition reduces to $P = (n-1)e + H + F$. This punishment exceeds that for the case when the whistle blower is a rat, $P = (n-1)e + F$, and corresponds to proportionality when $H = e - F$, the prescribed treatment of whistle blowers who seek to avoid guilt by association under SOX. Further, in comparison with the punishment, $P = (n-1)e + H + F$, the 'carrot' for reporting, H, is smaller than the 'stick' for violating, P.

Yet SOX contains no provisions for public policy claims by whistle blowers, nor does it provide them with an avenue for seeking punitive damages. This is puzzling given the Congressional impetus for passing SOX. Moreover, under the public goods motivation, $\lambda_M = e/(P+e-F) = e/(H+ne)$ and $\lambda_V = (R+m)/(H+R+x) = m/(H+ne)$. As the hero's treatment of the whistle blower increases, λ_M and λ_V decrease, thereby moving toward the ideal of (N, A).

RESULT 5: Sarbanes-Oxley mandates a no-retaliation policy for whistle blowers, $R = 0$. It does not; however, contain adequate provisions for whistle blower compensation. Increasing this compensation and/or the punishment for a violation approaches the ideal of (N, V).

Given that SOX does not preempt state and federal law, legal analysts have predicted that SOX whistle blowers are likely to make multiple claims − including public policy and state statutory claims − in conjunction with a claim under SOX. Our analysis supports increasing the range of whistle blower's claims, but it also suggests that whistle blower's compensation, $H = e - F$, should bounded from above by the violator's gain from wrongdoing, e, because of the potential for false accusations, F.

VI. Conclusion

In their 2002 *Time Magazine* 'Persons of the Year' joint interview, Cynthia Cooper, Coleen Rowley, and Sherron Watkins avoided being called whistle blowers. In popular culture the term is often pejorative; associated with tattletales, disloyal employees or individuals seeking self-aggrandizement and/or revenge. Further, there is a high personal cost associated with whistle blowing that too often includes job loss, retaliation and career blacklisting. Our analysis confirms the concerns of these and other whistle blowers by examining the effects of organizational culture on the calculus of ethical decision making.

This paper presents a model of organizational culture that accounts for the lack of hierarchy and role conflict involved in the whistle blowing decision. We examine the nexus of ethical rationales for whistle blowing (avoiding guilt by association versus providing a public good) and organizational attitudes toward whistle blowing (reducing type II errors, whistle blower as 'rat,' and provisions in Sarbanes Oxley). When these considerations are addressed within an evolutionary model of organizational culture and role duality, the resulting equilibrium

condition can be used to characterize the proportionality of punishment, an attribute that is largely absent in prior investigations of inspection games. For example, norms prescribing that the punishment must fit the crime (akin to *lex talionis*) are closely associated with incentives for whistle blowing that recognize the need to avoid type II errors (false accusations). In this case violations are minimized when whistle blowers are treated as heroes.

Within this context we find that the Sarbanes-Oxley Act is geared toward reducing type II errors, but it inadequately provides sufficient incentives for whistle blowing. In particular, whistle blowers should be rewarded for serving the public good. Again, the whistle blower is a hero. Operationalizing this recommendation has the added benefit of creating proportional rules of thumb for penalizing violators. These findings are especially important in an era where compliance with Sarbanes-Oxley dominates every major US corporate agenda and Congress is revisiting the law to address its unintended consequences.

Appendix: proofs.

RESULT 1: let $E_r[\lambda_r, \lambda_c]$ and $E_c[\lambda_r, \lambda_c]$ be the expected payoff for players in the row (r) and column (c) roles in Box 1, respectively, when facing (possibly mixed) row strategy λ_r and column strategy λ_c. In a mixed strategy equilibrium, $E_r[M, \lambda_c] = E_r[N, \lambda_c]$; i.e., $[H–m]\lambda_V – [R+m]\lambda_A = -x\lambda_V$. Given $\lambda_A = 1 - \lambda_V$, this implies $\lambda_V = (R+m)/(H + R + x)$. In the same way, $E_c[\lambda_r, V] = E_c[\lambda_r, A]$ implies $-P\lambda_M + e\lambda_N = -F\lambda_M$. As $\lambda_N = 1 - \lambda_M$, the equality simplifies to $\lambda_M = e/(e + P – F)$. Given that the left-hand and right-hand subgames in Figure 1 are identical, and equivalent to Box 1, this completes the proof. ■

References

Andreozzi, Luciano. 2004. Rewarding Policemen Increases Crime. Another Surprising Result From the Inspection Game. *Public Choice* 121: 69-82.

Arce, Daniel G. 2004. Conspicuous By Its Absence: Ethics and Managerial Economics. *Journal of Business Ethics* 54: 259- 5.

Arce, Daniel G. 2007. Is Agency Theory Self-Activating? *Economic Inquiry*. OnlineEarly Articles: doi:10.1111/j.1465-7295.2007.00047.x

Becker, Gary S. (1968). Crime and Punishment. An Economic Approach. *Journal of Political Economy* 76: 169-216.

Bowie, Norman. 1982. *Business Ethics*. Englewood Cliffs, NJ Prentice-Hall.

Brams, Steven J., and D. Marc Kilgour. 1985. Optimal deterrence. *Social Philosophy and Policy* 3: 118-35.

Brams, Steven J., and D. Marc Kilgour. 1987. Optimal threats. *Operations Research* 35: 524-36.

Brewer, Gene A., and Sally Coleman Selden. 1998. Whistle Blowers in the Federal Civil Service: New Evidence of the Public Service Ethic. *Journal of Public Administration Research and Theory* 8: 413-439.

Casson, Mark. 1991. *The Economics of Business Culture. Game Theory, Transactions Costs, and Economic Performance*. Oxford: Oxford University Press.

Ellingsen, Tore 1997. The Evolution of Bargaining Behavior. *Quarterly Journal of Economics*. 112: 581-602.

Fudenberg, Drew and Eric Maskin 1990. Evolution and Cooperation in Noisy Repeated Games. *American Economic Review*, *Papers and Proceedings*, 80: 274-9.

Gundlach, M.J., Douglas, S.C., and M.J. Martinko. 2003. The Decision to Blow the Whistle: A Social Information Processing Framework. *Academy of Management Review* 18: 107-123.

Heyes, Anthony G., 2005. Whistleblowers and the Regulation of Environmental Risk. Working paper, Royal Holloway, University of London.

Hirshleifer, Jack. 1977. Economics from a Biological Viewpoint. *Journal of Law & Economics* 20: 1-52.

Hofbauer, Josef and Karl Sigmund 1998. *Evolutionary Games and Population Dynamics*. Cambridge: Cambridge University Press.

Jubb, Peter. 1999. Whistleblowing: A Restrictive Definition and Interpretation. *Journal of Business Ethics* 21: 77-94.

Koch, Charles S. (2007). *The Science of Success*. Hoboken, NJ: Wiley.

Kolstad, Ivar 2007. The Evolution of Social Norms: With Managerial Implications. *Journal of Socio-Economics* 36: 59-72.

Maynard Smith, John. 1982. *Evolution and the Theory of Games*. Cambridge: Cambridge University Press.

Noreen, Eric 1988. The Economics of Ethics: A New Perspective on Agency Theory. *Accounting, Organizations and Society* 13: 359-69.

Rapoport, Anatol. 1994. Game Theory without Rationality. *The Behavioral and Brain Sciences* 7: 114-115.

Selten, Reinhard. 1980. A Note on the Evolutionary Stable Strategies in Animal Conflicts. *Journal of Theoretical Biology* 84: 93-101.

Time Magazine. 2002. Persons of the Year 2002. 30 December.

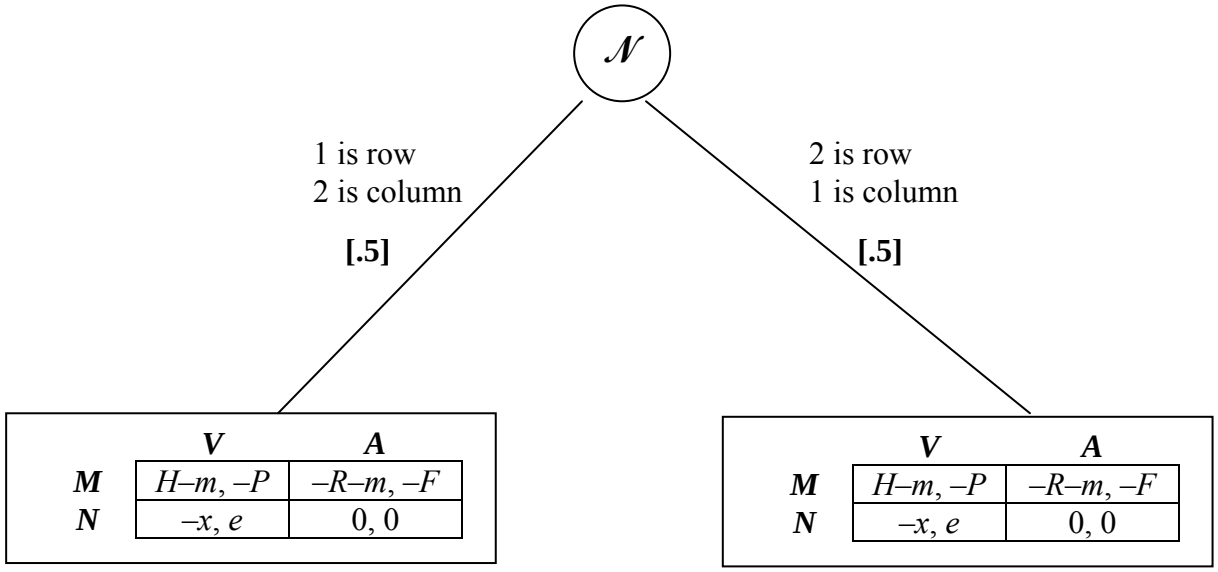
von Wangenheim, Georg 2004. *Games and Public Adminstration. The Law and Economics of Regulation and Licensing*. Cheltenham, UK: Edward Elgar.

Box 1: Types and Payoffs
[Role-Contingent Payoffs from Pairwise Encounters]

Roles: ↓Monitor/Worker→	**Violate (*V*)**	**Adhere (*A*)**
Monitor (*M*)	$H–m, -P$	$–R–m, –F$
Not (*N*)	$–x, e$	0, 0

Note: $H, m, P, R, F, e, x > 0$; $H - m > –x$; $–F > –P$.
$H \equiv$ degree to which whistle blower is treated as a 'hero;'
$m \equiv$ cost of monitoring;
$P \equiv$ punishment for an ethical breach and/or guilt by association;
$R \equiv$ degree to which whistle blower is treated like a 'rat;'
$F \equiv$ consequences for facing false accusations;
$x \equiv$ consequences for not whistle blowing; and
$e \equiv$ degree/value of ethical breach.

Figure 1: Cohort Duality with NonAssortative Matching
[Selten's (1980) Asymmetric Transformation]



Box 2: Asymmetric Contest for Box 1
[Role-Contingent Strategies, Nonassortative Matching]

	MV	***MA***	***NV***	***NA***
MV	$H–P–m$, $H–P–m$	$–(R+P+m)$, $H–F–m$	$H+e–m$, $–(e+P)$	$e–(R+m)$, $–(F+x)$
MA	$H–F–m$, $–(R+P+m)$	$–(R+F+m)$, $–(R+F+m)$	$H–m$, $–P$	$–(R+m)$, $–F$
NV	$–(P+x)$, $H+e–m$	$–P$, $H–m$	$e–x$, $e–x$,	e, $–x$
NA	$–(F+x)$, $e–R–m$	$–F$, $–(R+m)$	$–x$, e	0, 0

Table 1: Motives for Whistle Blowing and the Proportionality of Punishment

$P \equiv$ punishment. $\lambda_M \equiv$ frequency of monitoring. $\lambda_V \equiv$ frequency of violations.		**Whistle Blowing Ethic**	
		Guilt By Association $x = P$	**Public Good $x = ne$**
Corporate Culture	**Whistle Blower is a Rat: $H = -R$**	$P > e$ $\lambda_M = e/P$; $\lambda_V = (R+m)/P$	$P = (n-1)e + F$ $\lambda_M = 1/n$; $\lambda_V = (R+m)/ne$
	Type II	When $H = P - (R + F)$, $P = e$ (proportionality)	When $H = e - (R + F)$**,** $P = ne$ (proportionality)
	Sarbanes Oxley $R = 0$	P unspecified $H = e - F$ $\lambda_M = e/(P+H)$; $\lambda_V = (R+m)/ (P+H)$	$P = (n-1)e + F + H$ $\lambda_M = e/(P+ne)$; $\lambda_V = (R+m)/(P+ne)$