Whistleblowing Behavior in Organizations

Quinteros, María José and Villena, Marcelo J. and Villena, Mauricio G.

Universidad Finis Terrae, Universidad Adolfo Ibáñez, Universidad Adolfo Ibáñez

24 November 2019

Whistleblowing Behavior in Organizations

María J. Quinteros[1,2] Marcelo J. Villena,[2] Mauricio G. Villena,[3]

[1] Universidad Finis Terrae, Business School, Chile
[2] Universidad Adolfo Ibáñez, Engineering School, Chile
[3] Universidad Adolfo Ibáñez, Business School, Chile

November 24, 2019

Abstract

We develop a game theoretical model of whistleblowing behavior in organizations, focusing specifically on the role of incentives aimed at encouraging this type of behavior. We also analyze the potential impacts of whistleblowing behavior on the persistence of corruption. First, we present a static game consisting of two employees with three available strategies: honest, corrupt and whistleblowing behavior. Later, we examine the pure and mixed Nash equilibrium strategies of the game. Second, we use the concept of replicator dynamics to formally explore the local asymptotic stability of whistleblowing behavior within organizations. Our main results show that whistleblowing as a mechanism to control wrongdoing is only relevant under the existence of external monitoring (if the probability of detecting wrongdoing with an external mechanism is close to zero, then in the long term, all employees will begin to behave corruptly). We also show that whistleblowers reduce the minimum wages required to avoid corruption within an organization, making it less costly for an organization to combat corruption. Finally, we claim that whistleblowing strategies seem to be less attractive for activities with very high bribery in comparison to the rewards for whistleblowers, for example, this could be the case of manufacturing or retail, but not for financial services in general.

Keywords: whistleblowing, corruption, game theory, replicator dynamics, incentives.

1 Introduction

Corruption is a persistent feature in human societies, and similar to crime in general, it has always existed. Thus, corruption has always been an important concern for managerial teams in all kinds of organizations. Interestingly, most studies that have modeled corruption have focused on countries and governments, as well as on the use of punishment and surveillance to increase the expected cost of crime; see (Rose-Ackerman, 1975) for an early work in this topic and (Aidt, 2003) for a lucid survey of the economic literature.

Recently, whistleblowing has become an important monitoring mechanism in the wake of numerous corporate scandals involving accounting firms (Alleyne, Hudaib, & Pike, 2013). The business community and interested regulatory agencies are now calling for whistleblowing to be a prominent part of organizational culture.

For instance, in 2010, the U.S. Securities and Exchange Commission (SEC) established the Dodd-Frank Whistleblower Program. The Dodd-Frank Act of 2010 included Section 21F ("Securities Whistleblower Incentives and Protections"), which directs the SEC to issue monetary awards to individuals who voluntarily provide original information leading to successful enforcement of monetary sanctions of over $1 million. During 2017 alone, the SEC has ordered offenders to pay more than $975 million in total monetary sanctions in enforcement matters involving whistleblower information; this amount includes more than $671 million in disgorgement of profits acquired through securities fraud, which are in turn returned to investors who have been harmed. Whistleblowers are typically compensated in an amount equal to 10 to 30 percent of the monetary sanctions collected. In addition to establishing a reward program to encourage the submission of high-quality information, the SEC has implemented regulations prohibiting retaliation by firms against employees who have reported possible wrongdoing based on a reasonable belief that a possible securities violation has occurred, is in progress, or is about to occur.

Following the collapse of Arthur Andersen in the Enron scandal, whistleblowing has received considerable attention, primarily from the ethics literature; since then, business research on whistleblowing has also started to expand (Alleyne et al., 2013). Nevertheless, from a management science perspective, research on whistleblowing behavior has been rather scarce. In this context, this study aims to contribute by formally investigating the role of whistleblowing in and potential impact of whistleblowing on the persistence of corruption in organizations, focusing specifically on the role of incentives aimed at encouraging whistleblowing behavior.

Whistleblowing can be formally defined as the disclosure by organization members (former or current) of illegal, immoral, or illegitimate practices under the control of their employers to persons or organizations that may be able to effect action (Near & Miceli, 2016). A recent example of whistleblowing behavior is that of a Monsanto executive who reported to the SEC improper accounting at the genetic engineering and pesticides producer Agrocorp. In an article on the subject, the New York Times reported that "*the whistleblower got stymied within the company*" and incurred many expenses since he had to fly to the U.S. Securities and Exchange Commission headquarters in Washington D.C. many times to assist them and that he finally had to leave Monsanto during the course of the investigation[1]. The whistleblower was reported to have stated that: "*It's really difficult when your company is doing something you know is wrong but you've got everybody around you saying it's perfectly fine... the Monsanto culture is very tightknit. Everybody has stock options and everyone is financially at risk. So they go with the flow*". After the SEC was informed that the company's accounting practices for its Roundup herbicides were being used to overstate earnings, Monsanto agreed to settle for $80 million, and the Monsanto executive earned a $22 million award for blowing the whistle. While this case illustrates accusations of wrongdoing by an employee to her organization, there are few cases where an employee blows the whistle on wrongdoing by another employee.

From this business case, we can infer some distinctive elements about how whistleblowing functions in practice. For example, spontaneous whistleblowing behavior can be difficult to initiate in practice because it implies costly activities and typically garners no economic rewards (Heyes & Kapur, 2008). The costs of whistleblowing include monitoring and transaction costs associated with actually reporting corrupt behavior. These costs clearly depend on the number of agents within an organization that somehow favor the corrupt activities being committed. Indeed, the monitoring and transaction costs of reporting illicit activities can increase when more people within an organization support this type of behavior. Thus, one way to approach the problem is to model corruption as a social norm of behavior assuming that the illicit activities are somehow accepted by most agents as valid and common practices so that going against them means going against the majority.

In this study, we present a theoretical model for approaching corruption, focusing specifically on the role of incentives aimed at encouraging whistleblowing behavior[2]. We model corruption as a social norm of behavior using elements of a classic game theory approach and elements of evolutionary game theory (EGT). For examples, see (Cai & Kock, 2009), (Liu, Liao, & Wei, 2015), (Cheng, Bai, & Yang, 2017), (Anastasopoulos & Anastasopoulos, 2012), (Hellmann & Staudigl, 2014). Specifically, we use the concept of replicator dynamics to formally explore the local asymptotic stability of the following types of behavior within an organization: (i) honest behavior, which implies that an employee does not receive any bribes from illicit activity; (ii) corrupt behavior, in which a member of the organization does receive bribes from a corrupt relationship; and (iii) whistleblowing (honest enforcer) behavior, in which an employee not only behaves honestly, i.e., does not receive bribes, but also monitors other people within the organization and reports them if they are behaving corruptly. Assuming that an organization includes three different population shares pursuing these three different behavior types, we use the concept of replicator dynamics to analyze which of these population shares will become stable within the organization in the long run. In other words, we formally explore the asymptotic stability of the noncorruption equilibrium, if all individuals behave honestly, and the corruption equilibrium, if all individuals within the organization behave corruptly.

In particular, our work adds to the variables of analysis related to the environment for whistle-

[1]See https://www.nytimes.com/2016/09/11/business/for-monsanto-whistleblower-a-22-million-award-that-fell-short.html

[2]For an early and less complete version of this paper, see (Villena & Villena, 2010).

blowing behavior within organizations in order to provide guidance on how economic instruments aiming to promote whistleblowing behavior should be designed to effectively help break the stability of corruption. To the best of our knowledge, this work is one of the first studies to specifically address the impact of economic incentives for whistleblowers and the stability of corruption. Consequently, in this article, we study what would happen to the stability of corruption with these instruments in place versus a situation in which they are not. Furthermore, we analyze the effects of economic incentives on the stability of corruption.

The remainder of this paper is organized as follows. Section 2 takes us through the literature review, section 3 describes the problem of interest, section 4 presents the static game of whistleblowing behavior, section 5 shows the evolution of whistleblowing behavior using replicator dynamics, and finally section 6 presents some concluding remarks and notes potential topics for future research.

2 Literature review

As noted above, most studies on modeling corruption have focused on countries and governments, with the costs and benefits of crime almost being treated like those of a working activity. In one of the first works on this topic, (Rose-Ackerman, 1975) considered the relationship between market structure and the incidence of corrupt dealings in the context of government contracting processes. She noted that it is possible to consider the extent to which various criminal sanctions will deter corruption by revising contracting procedures. The work of (Rose-Ackerman, 1975) and the following works on corruption in economics, criminology and even sociology have doubtlessly been influenced decisively by the seminal paper of (Becker, 1968).

Indeed, Becker introduced the modeling of individual criminal behavior as a response to incentives. An agent decides whether to commit a crime–and how much crime to commit–by comparing the benefits and costs of crime with those of alternative activities. In Becker's paper and its successors, (Ehrlich, 1973), (Block & Heineke, 1975), and (Freeman, 1996), among others, the aim is to study to a certain extent how the probability of being caught, the magnitude of the punishment, the proceeds of criminal activity and the return to work (alternative to crime) would affect the level of crime. However, the Beckerian model faces agent decisions at a moment in time where only external control methods exist.

An extension to the literature on crime is the introduction of internal auditing of wrongdoing, for example, through whistleblowers. In this context, some theoretical works have explored whistleblowing behavior from an economic perspective. (Acemoglu & Jackson, 2017) study how social norms can constrain the effectiveness of laws that ban certain types of behavior when laws are in conflict with social norms, with the result being that compliance and enforcement are weaker.

(Arce, 2010) presents a model based on game theory that elaborates on the expectations and conviction that is symptomatic of a whistleblower as they are preparing for the act of exposing what they believe to be wrongdoing.

(Søreide, 2008) presents an economic framework aimed at explaining the potential reaction of multinationals to the loss of a contract because a competitor has offered a bribe.

(Heyes & Kapur, 2008) puts forward a general whistleblower "motivation function", which relates individuals' propensity to blow the whistle to the characteristics of the observed malfeasance and the enforcement environment, characterizing the optimal policy for three specific cases.

Our work is also related to those of (Lui, 1986), (Cadot, 1987), (Sah & Stiglitz, 1988), (Andvig & Moene, 1990), (Murphy, Shleifer, & Vishny, 1991), (Acemoglu, 1995), (Baker, Gibbons, & Murphy, 1994) and (Tirole, 1996) in the sense that we also emphasize the self-reinforcing nature of corruption, implying that the greater the number of people who adhere to corrupt activities is, the more persistent corruption becomes.

Including whistleblowing behavior in the analysis allows us to consider additional control instruments that an organization can use to prevent corruption, going beyond the usual variables mentioned in the economic literature, namely, the employees' wage rate, the external monitoring system, and the penalties for corrupt activities. Some specific questions we attempt to answer in this context are the following: How can the number of whistleblowers in an organization with an initially small population of employees who are willing to monitor and report corrupt behavior be increased? Furthermore, when corruption is already in place as a social norm of behavior in an organization, how can whistleblowing policies help break the stability of corruption?

3 The model

We consider an organization in which employees display the following types of behavior:

1. Honest.
2. Corrupt.
3. Whistleblowing.

An employee receives a salary of w for her work. Therefore, an honest employee gets paid w. If an employee commits illicit acts, she receives an amount β for her illicit acts. However, if the corrupt employee is caught, she is fired, and she can get an alternative job a wage of $w_0 \geq 0$ and will pay to the actual firm a penalty of $f \geq 0$. The firm may have some whistleblower employees who monitor others and denounce them when caught in illicit acts. Whistleblowers can detect wrongdoing behavior with a probability of $\Theta > 0$. However, as discussed in the introduction, it is difficult for whistleblowing behavior to arise spontaneously since it is a costly activity. The costs associated with whistleblowing include monitoring costs of $m > 0$ and transaction costs of $\tau > 0$ associated with actually reporting wrongdoing.

We assume that whistleblowing employees are honest. Given that, it is clear that no payoff maximizer agent will become a whistleblower since this activity only lowers her payoff regardless of the behavior of the rest of the employees. In fact, in this case, the whistleblowing strategy is strictly dominated by the honest strategy since $w > w - \tau\Theta - m$, so no employee has the incentive to become a whistleblower. This, in turn, implies that no employee is deterred from corruption by the threat of whistleblowing behavior from her colleagues.

Because whistleblowing behavior is a dominating strategy, we introduce a reward denoted by $\sigma \geq 0$ for monitoring the work done by their colleagues and reporting illicit acts. In addition to whistleblowing behavior among employees, an external mechanism–for example, auditing–can also detect wrongdoing behavior with probability $\theta \geq 0$.

Considering a total population of n employees, a proportion of p_1 is honest, p_2 is corrupt and p_3 is made up of whistleblowers ($p_1 + p_2 + p_3 = 1$). Each of the $p_1 n$ honest employees receives the payoff associated with honest behavior, a wage of $\pi_1 = w$. The proportion of corrupt employees $p_2 n$ receive the expected payoff associated with wrongdoing activities, $\pi_2 = (1 - \theta - p_3\Theta)(w + \beta) + (\theta + p_3\Theta)(w_0 - f)$, assuming that corrupt acts can be caught by the external agency with probability θ or by any of the whistleblowers with probability $p_3\Theta$. If she is not caught, she receives her salary and the bribery payment, and if she is caught, she is fired and must work in another place and pay the penalty. Finally, each of the $p_3 n$ whistleblowers receives the expected payoff of $\pi_3 = w + p_2\Theta(\sigma - \tau) - m$, which is associated with her salary plus the reward for blowing the whistle less the costs associated with monitoring and reporting. The average payoff of all employees is $\bar{\pi} = p_1\pi_1 + p_2\pi_2 + p_3\pi_3$.

To study this problem, we first propose a two-player non-cooperative game; subsequently, we solve the same problem using an EGT approach.

4 The intuition of whistleblowing behavior

The effect of whistleblowing as a mechanism to control corruption within organizations can be first modeled using non-cooperative game theory. Formally, the n-player game specifies the players' strategy spaces $S_1, ..., S_n$ and their payoff functions $\pi_1, ..., \pi_n$. We denote this game by $G = \{S_1, ..., S_n; \pi_1, ..., \pi_n\}$.

Our game consists of two players, both employees of the firm, denoted by player 1 and player 2. Each player in this game has three available strategies: to behave honestly, to behave corruptly or to blow the whistle, represented by $S = \{H, C, WB\}$. Table 1 represents the payoffs of the game. Therefore, we define the game by $G = \{H, C, WB; \pi_H, \pi_C, \pi_{WB}\}$, where $\pi_H = (\pi_{hh}, \pi_{hc}, \pi_{hw})$, $\pi_C = (\pi_{ch}, \pi_{cc}, \pi_{cw})$ and $\pi_{WC} = (\pi_{wh}, \pi_{wc}, \pi_{ww})$.

Table 1: Whistleblowing two-player symmetric game

		Player 2 H	C	WB
Player 1	H	π_{hh}, π_{hh}	π_{hc}, π_{ch}	$\pi_{hw}, \pi_{wh},$
	C	π_{ch}, π_{hc}	π_{cc}, π_{cc}	π_{cw}, π_{wc}
	WB	π_{wh}, π_{hw}	$\pi_{wc}, \pi_{cw},$	$\pi_{ww}, \pi_{ww},$

Where:
$\pi_{hh} = \pi_{hc} = \pi_{hw} = \pi_h = w$
$\pi_{wh} = \pi_{ww} = w - m$
$\pi_{wc} = \Theta(\sigma - \tau) + w - m$
$\pi_{ch} = \pi_{cc} = (1 - \theta)(\beta + w) + \theta(w_0 - f)$
$\pi_{cw} = (1 - \theta - \Theta)(w + \beta) + (\theta + \Theta)(w_0 - f)$

Proposition 1 presents the equilibrium analysis of game G.

Proposition 1: Considering the two-player whistleblowing game G described above, we obtain the following results:

1. There is a unique Nash equilibrium given by the corrupt behavior of both employees {C,C} whenever $\pi_{cc} \geq \pi_{wc}$.
2. There is a unique Nash equilibrium given by the behavior of one of the employees being corrupt and the other being a whistleblower {(C,W),(W,C)} whenever $\pi_{cc} < \pi_{wc}$ and $\pi_{hw} < \pi_{cw}$.
3. The game has a single mixed-strategy equilibrium whenever $\pi_{cc} < \pi_{wc}$ and $\pi_{hw} \geq \pi_{cw}$.

Proposition 1 shows the existence of a mixed strategy for employees 1 and 2. Three outcomes are possible depending on the values of the parameters. The game may have a pure-strategy Nash equilibrium defined by an environment of corruption when the conditions in 1 are satisfied. The game may also present a pure-strategy Nash equilibrium of corrupt employees and whistleblowers when the conditions in 2 are satisfied. Finally, when the conditions in 3 are satisfied, a mixed strategy is reached by player i of behaving corruptly when the other player behaves honestly, behaving as a whistleblower when the other player is corrupt, or being honest when the other player is a whistleblower.

Proposition 1.1 implies that $w \leq \frac{\beta}{\theta} - \beta + \frac{m - \Theta(\sigma - \tau)}{\theta} - f - w_0$, which tells us that if an external enforcing agency does not exist, any organization will have corrupt employees regardless of the existence of whistleblowers. If we assume that $\theta > 0$, then the role of whistleblowers becomes relevant since the condition for minimum wages required to avoid corruption within an organization is reduced by $\frac{\Theta(\sigma - \tau)}{\theta}$.

The conditions in Proposition 1.2 imply that $\frac{\beta}{\theta} + \frac{m - \Theta(\sigma - \tau)}{\theta} < w < \frac{\beta}{\theta + \Theta}$. Since θ and Θ

$\in [0,1]$, $\frac{\beta}{\theta} > \frac{\beta}{\theta+\Theta}$, which implies that in order to be true, our inequality must be the case where $\Theta(\sigma - \tau)$ is high enough. Interpreting the result, Θ must have a minimum feasible value and the reward for blowing the whistle σ must be high.

When the conditions in Proposition 1.3 are met, there is no pure-strategy equilibrium for this particular game. However, we can find a mixed strategy for employee i probability distribution of her set of pure strategies (Hofbauer & Weibull, 1996). Let x_i denote the probability assigned by an employee to pure strategy $i \in S$, where $S = \{H, C, WB\}$. It should be noted that $x_i \in [0,1]$ and $\sum_{i \in S} x_i = 1$. The expected payoff of an employee is given by:

$$\pi(x) = x_1\pi_{hh} + x_2[(x_1 + x_2)\pi_{ch} + (1 - x_1 - x_2)\pi_{cw}] + (1 - x_1 - x_2)[(1 - x_2)\pi_{wh} + x_2\pi_{wc}] \quad (1)$$

Thus, the employee chooses x_1 and x_2 that maximize equation 1. Solving the first-order conditions, $\frac{\partial \pi}{\partial x_i} = 0$, $i = 1, 2$, the mixed-strategy equilibrium is:

$$\begin{bmatrix} x_1^* \\ x_2^* \\ x_3^* \end{bmatrix} = \begin{bmatrix} 1 + \frac{\theta(\beta+f+w-w_0)-\beta+m}{\Theta(\beta+f-\sigma+\tau+w-w_0)} \\ \frac{-m}{\Theta(\beta+f-\sigma+\tau+w-w_0)} \\ \frac{\theta(w_0-w-f-\beta)+\beta}{\Theta(\beta+f-\sigma+\tau+w-w_0)} \end{bmatrix} \quad (2)$$

To reach this mixed strategy, p_i must satisfy the conditions of probability, which means that $\frac{\beta}{\theta} + w_0 - f - \beta \leq w \leq \sigma + w_0 - \beta - f - \tau$, which is true when $\sigma - \tau \geq \frac{\beta}{\theta}$.

We recover the Beckerian case when there is only an external mechanism that detects wrongdoing, with a unique Nash equilibrium given by corrupt employees if $w < \frac{\beta}{\theta} + w_0 - \beta - f$ and a unique Nash equilibrium given by only honest employees otherwise, meaning that the lower the probability of catching wrongdoing is, the higher the bribe is, the higher the alternative salaries are or the lower the penalty when a corrupt employee is caught is, the higher the wages that must be paid to avoid corruption.

We are interested in studying what happens with our set of optimal solutions (x_1^*, x_2^*, x_3^*) when the value of any of the parameters changes, namely, $\frac{\partial x_i^*}{\partial a}$. Comparative statistical results are presented in table 2, showing whether the value of the optimal probability increases or not when there are changes in the parameters that define our problem.

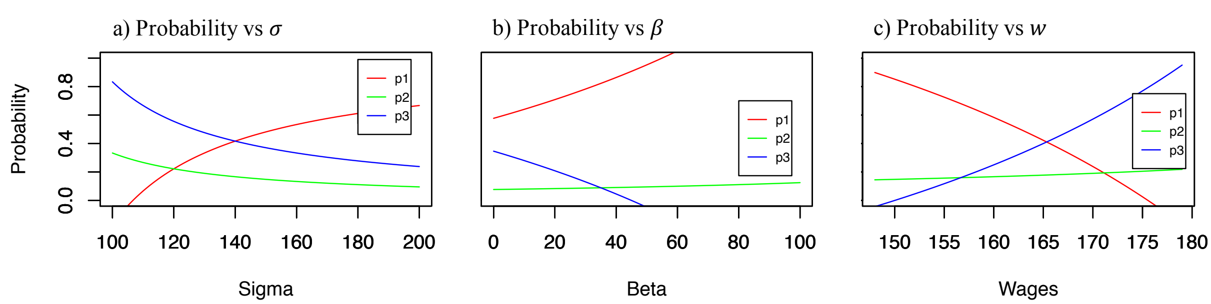
Table 2: Results of the comparative statistics

	x_1	x_2	x_3
σ	$+$	$-$	$-$
β	$+$	$+$	$-$
w	$-$	$+$	$+$
θ	$-$	no effect	$+$
Θ	$+$	$-$	$-$

An increase in reward σ implies that the incentives for wrongdoing are less attractive, so a percentage of corrupt employees will prefer to start behaving honestly. At the same time, as there is a smaller proportion of corrupt employees, it is less attractive to behave as whistleblowers, also decreasing the proportion of whistleblowers. If the bribery β increases, behaving corruptly is more attractive than being a whistleblower; therefore, a proportion of whistleblowers begin behaving corruptly, and another proportion of whistleblowers migrate toward behaving honestly. The introduction of whistleblowers implies a counterintuitive result for salaries since an increase in w leads to an increase in the corrupt and whistleblowing population. If the external probability of catching wrongdoing θ increases, only honest employees move toward whistleblowing, which has no effect on the proportion of employees who behave corruptly. Finally, an increase in Θ has the same effect as an increase in the reward: it increases honest behavior among employees and decreases corrupt and whistleblowing behavior.

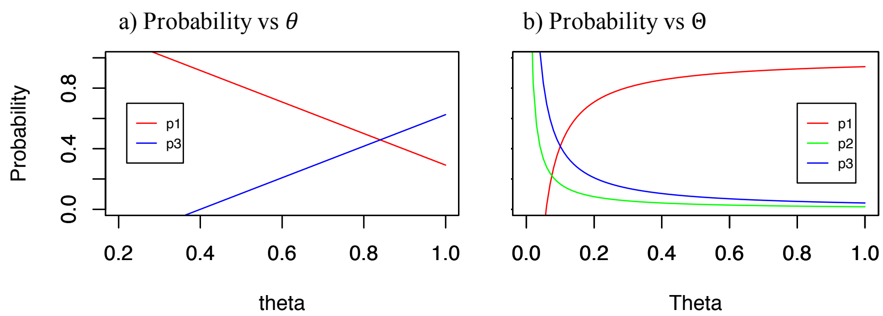
Apart from the comparative statistical results shown in table 2, we now study the magnitude of those effects. To do so, we graphically illustrate what happens to the probabilities when each parameter is varied. Figure 1 shows the movement of probabilities as a function of σ, β and w. Panel a) shows that increasing the reward for whistleblowing increases the number of honest employees in the long term while decreasing the number of corrupt and whistleblower employees. Therefore, the reward becomes an instrument that the firm announces but only actually pays a few times. Panel b) shows that the bribery β, as expected, increases the probability of behaving corruptly but also increases the probability of being honest, thereby decreasing the probability of behaving as a whistleblower. Panel c) shows the change in the probabilities given a change in salary. The results are counterintuitive, as a change in salary increases the probability of behaving as a whistleblower but also the probability of behaving corruptly, although the rate of this change is less than the increase in the number of whistleblowers, and decreases the probability of employees behaving honestly. Figure 2 shows how the probability of being detected

Figure 1: Change in the probability of behaving honestly, corruptly or as a whistleblower as a function of parameters σ, β and wages



impacts the probabilities of being honest, corrupt or a whistleblower. Panel a) shows that when the probability of an external agency catching corruption increases, the probability of being a whistleblower increases, and the probability of being honest decreases; however, the probability of being corrupt is not a function of θ. Panel b) shows that an increase in the probability of an employee catching wrongdoing implies a decrease in the probability of behaving corruptly or whistleblowing.

Figure 2: Change in the probability of behaving honestly, corruptly or as a whistleblower as a function of the probability of catching corruption



5 The dynamics of whistleblowing behavior

We now model whistleblowing behavior within an organization with the replicator dynamic equation, formulating a simple model of evolution and biased learning in games in which successful strategies spread by natural selection among a population.

To formalize the replicator dynamic equation, consider an evolutionary game with q pure strategies and stage game payoff π_{ij} for any i-player who meets any j-player. If $p = (p_1, ..., p_q)$ is the frequency of each type of population, the expected payoff for the i-player is then $\pi_i(p) = \sum_{j=1}^{q} p_j \pi_{ij}$, and the average payoff in the game is $\bar{\pi}(p) = \sum_{j=1}^{q} p_j \pi_i$. The replicator dynamic game for our model is then given by:

$$\frac{\partial p_i}{\partial t} = \dot{p_i} = p_i(\pi_i(p) - \bar{\pi}(p)) \tag{3}$$

The replicator equation 3 expresses the idea that strategies grow among population if they do better than average. The strategies that do best grow fastest. It is clear that a Nash equilibrium is a stationary point in the dynamic system. Conversely, each stable stationary point is a Nash equilibrium, and an asymptotically stable fixed point is a perfect equilibrium.

The replicator dynamics of our model are represented by the two differential equations presented below:

$$\dot{p_1} = p_1(\pi_1 - \bar{\pi}) \tag{4}$$

$$\dot{p_2} = p_2(\pi_2 - \bar{\pi}) \tag{5}$$

The rate of growth of the share of the population using strategies 1 and 2 are proportional to the amount by which those strategies' payoffs exceed the average payoff of the strategies in the population.

We are interested in examining the stability of the system since stability is what tells us where the model converges under certain given conditions to find the conditions that encourage employees to behave honestly. Our model is nonlinear, and we use the (indirect) Lyapunov's linearization method to study the stability of each fixed point. Let p_1^* and p_2^* be the fixed points of the original system, then:

$$\dot{p_1} = p_1 \frac{\partial \dot{p_1}(p_1^*, p_2^*)}{\partial p_1} + p_2 \frac{\partial \dot{p_1}(p_1^*, p_2^*)}{\partial p_2} \tag{6}$$

$$\dot{p_2} = p_1 \frac{\partial \dot{p_2}(p_1^*, p_2^*)}{\partial p_1} + p_2 \frac{\partial \dot{p_2}(p_1^*, p_2^*)}{\partial p_2} \tag{7}$$

To study stability, we only need to find the eigenvalues of the system presented above, and if $Re(\lambda_i) < 0$, then (p_1^*, p_2^*) is asymptotically stable.

By solving the system $\dot{p_1} = 0$, $\dot{p_2} = 0$, we found 5 solutions in which the system converges (fixed points of the system). The first equilibrium is when all employees in the long term become whistleblowers, represented by $p_1 = p_2 = 0$ and $p_3 = 1$. A second equilibrium is when all employees in the long term become honest ($p_1 = 1$ and $p_2 = p_3 = 0$). The third equilibrium is when all employees in the long term are corrupt ($p_1 = p_3 = 0$ and $p_2 = 1$). The fourth case is when employees in the long term are either corrupt or whistleblowers, represented by $p_1 = 0$, $p_2 > 0$ and $p_3 > 0$. The last equilibrium is when the 3 types of employees coexist: honest employees, corrupt employees and whistleblowers ($p_1 > 0$, $p_2 > 0$ and $p_3 > 0$).

5.1 The stability of only whistleblower employees

By solving equations (6) and (7) with fixed points $p_1^* = p_2^* = 0$, the eigenvalues are $\lambda_1 = m$ and $\lambda_2 = -\Theta\beta - \Theta f - \Theta w + \Theta w_0 - \beta\theta - f\theta - \theta w + \theta w_0 + \beta + m$. It is clear that the solution $p_1^* = p_2^* = 0$ is unstable since $\lambda_1 = m > 0$, meaning that the solution is unstable regardless of the values of the other parameters.

Proposition 2: It is a dominant strategy to behave honestly over whistleblowing when there are no corrupt employees.

Proof: The payoff for an honest employee is $\pi_h = w$. The payoff for a whistleblower employee when there is no corrupt population is $\pi_w = w - m$. Since $w > w - m$, it is clear that when there is no wrongdoing, the population will all behave honestly. Proposition 2 makes sense since if no corruption is detected within the population , a whistleblower only incurs the costs of monitoring and never receives the reward.

5.2 The stability of the honest equilibrium

Proposition 3: The honest equilibrium ($p_1^* = 1$, $p_2^* = 0$) is locally asymptotically stable if and only if $w \geq w_0 - \beta - f + \frac{\beta}{\theta}$.

This is easy to show. The eigenvalues for $p_1^* = 1$, $p_2^* = 0$ are $\lambda_1 = -m$ and $\lambda_2 = -\beta\theta - f\theta - \theta w + \theta w_0 + \beta$. Since $\lambda_1 < 0$, we only need to impose the condition $\lambda_2 < 0$ in order to get a stable fixed point, which is true when $w \geq w_0 - \beta - f + \frac{\beta}{\theta}$. Thus, efficiency wages $w > w_0$ for employees, high penalty rates f and a high probability of being detected by an external agency θ imply significant expected costs of dismissal, so the honest equilibrium will be stable and hard to break. By contrast, the higher the amount of the bribery β is, the more easily our honest equilibrium can be broken.

However, the design of a successful external control and monitoring mechanism can be both difficult and costly. In practice, it may imply a very low θ. If $\theta \to 0$ implies that the efficiency wage w becomes too high and actually very hard to pay, the stability of the honest equilibrium is easy to break. This may imply that becoming a firm with only honest employees is never achieved given the importance of a potential equilibrium including a population of whistleblowers.

5.3 The stability of the corruption equilibrium

An environment of corruption occurs when $p_1^* = 0$, $p_2^* = 1$. In this case, the eigenvalues are $\lambda_1 = \beta\theta + f\theta + \theta w - w_0\theta - \beta$ and $\lambda_2 = \Theta\sigma - \Theta\tau + \beta\theta + f\theta + w\theta - w_0\theta - \beta - m$. The conditions for employees becoming corrupt in the long run depend on Θ.

Proposition 4: The corruption equilibrium is asymptotically stable if and only if:

1. Whenever $\Theta \leq \frac{m}{\sigma-\tau}$ and $w \leq w_0 - \beta - f + \frac{\beta}{\theta}$ or
2. Whenever $\Theta > \frac{m}{\sigma-\tau}$ and $w \leq w_0 - \beta - f + \frac{\beta}{\theta} + \frac{m}{\theta} - \Theta\frac{\sigma-\tau}{\theta}$.

The proof of proposition 4 is as follows. To obtain stability, we must impose $\lambda_1 < 0$ and $\lambda_2 < 0$. For λ_1, the condition is true if $w \leq w_0 - \beta - f + \frac{\beta}{\theta}$; however, for λ_2, we have 2 cases: 1) when $\Theta\sigma - \Theta\tau - m < 0$, $\lambda_1 < 0$ and $\lambda_2 < 0$ if $w \leq w_0 - \beta - f + \frac{\beta}{\theta}$. When $\Theta\sigma - \Theta\tau - m > 0$, we then must impose $\Theta\sigma - \Theta\tau + \beta\theta + f\theta + w\theta - w_0\theta - \beta - m < 0$, which is equivalent to $w \leq w_0 - \beta - f + \frac{\beta}{\theta} + \frac{m}{\theta} - \Theta\frac{\sigma-\tau}{\theta}$.

Assuming $\sigma > \tau$, when the probability of catching a corrupt employee is low, $\Theta < \frac{m}{\sigma-\tau}$, then wages must be high to avoid corruption. The wage condition is relaxed to some extent (by $\frac{m}{\theta} - \Theta\frac{\sigma-\tau}{\theta}$) when the probability $\Theta > \frac{m}{\sigma-\tau}$. If the probability of being detected by an external agency is rather small, the corruption equilibrium will be stable and hard to break. In fact, if $\theta \to 0$, the condition for the local asymptotic stability of the corruption equilibrium is always satisfied, which implies that whenever corruption becomes a common practice, it will be very difficult to stop. In the same way, if the amount of bribery, β, and the salary if dismissed, w_0, are high in comparison with w and the penalty for being caught in a corrupt activity, f is low and the corruption equilibrium will also be stable.

When $\sigma < \tau$, namely the benefit of whistleblowing, is less than the reporting costs, then it is always the case that $\Theta > \frac{m}{\sigma-\tau}$, the minimum salary for employees to not become corrupt increases. From the conditions, it becomes clear that if the probability of being detected by an external agency θ is rather small, the corruption equilibrium will be stable and hard to break.

By contrast, an increase in salaries may break the stability of corruption. This argument dates back to (Becker & Stigler, 1974) who noted that efficiency wages can be used to control corruption since they increase the cost of dismissal and, therefore, make employees more reluctant to accept bribes. Nevertheless, paying high salaries can be very expensive for firms and does not ensure that corruption will be reduced in all situations.

5.4 The stability of corrupt and whistleblowing employees

This means that in the long run, $p_1^* = 0$ and $p_2^* = b$, $b \in (0,1)$. In this environment, a firm has employees that are either corrupt or whistleblowers. In this case, the eigenvalues are given by $\lambda_1 = i - kw$ and $\lambda_2 = j - lw$, where:

- $i = -\Theta(f + \tau - w_0 + \beta - \sigma)b^2 + (\Theta(f + \tau - w_0 + \beta - \sigma) + (f - w_0 + \beta)\theta - m - \beta)b + m$
- $j = -3\Theta(f + \tau - w_0 + \beta - \sigma)b^2 + ((4f + 2\tau - 4w_0 + 4\beta - 2\sigma)\Theta + (2f - 2w_0 + 2\beta)\theta - 2m - 2\beta)b + (-f + w_0 - \beta)\Theta + (-f + w_0 - \beta)\theta + m + \beta$
- $k = b(\Theta b - \Theta - \theta)$
- $l = 3\Theta b^2 + (-4\Theta - 2\theta)b + \Theta + \theta$

Proposition 5: The equilibrium of corrupt employees and whistleblowers is asymptotically stable if and only if $\Theta > \frac{m}{\sigma - \tau}$ and $w > w_0 - \beta - f + \frac{\beta}{\theta} - \Theta\frac{\sigma - \tau}{\theta}$ and if:

1. $w > \frac{j}{l}$ or
2. $w < \frac{i}{k}$.

Proof of proposition 5: First notice that if $\Theta < \frac{m}{\sigma - \tau}$ the replicator dynamic is always asymptotically stable and converges to either all honest or all corrupt employees. If $w \leq w_0 - \beta - f + \frac{\beta}{\theta}$, the model converges to the corruption equilibrium when $\Theta < \frac{m}{\sigma - \tau}$ (from proposition 3). If $w \geq w_0 - \beta - f + \frac{\beta}{\theta}$, the model always converges to the honest equilibrium, in particular when $\Theta < \frac{m}{\sigma - \tau}$. If $\Theta > \frac{m}{\sigma - \tau}$ and $w < w_0 - \beta - f + \frac{\beta}{\theta} - \Theta\frac{\sigma - \tau}{\theta}$, then the model converges to the corruption equilibrium, so we only need to show stability of the corrupt employees and whistleblowers when $\Theta > \frac{m}{\sigma - \tau}$ and $w > w_0 - \beta - f + \frac{\beta}{\theta} - \Theta\frac{\sigma - \tau}{\theta}$. Notice that $k < 0$ and $l > 0$.

If $\lambda_1 < \lambda_2 \Rightarrow$, we need to impose $\lambda_2 < 0 \Rightarrow j - lw < 0 \Rightarrow w > \frac{j}{l}$. If $\lambda_2 < \lambda_1 \Rightarrow$, we need to impose $\lambda_1 < 0 \Rightarrow i - kw < 0 \Rightarrow w < \frac{i}{k}$ (remember that $k < 0$).

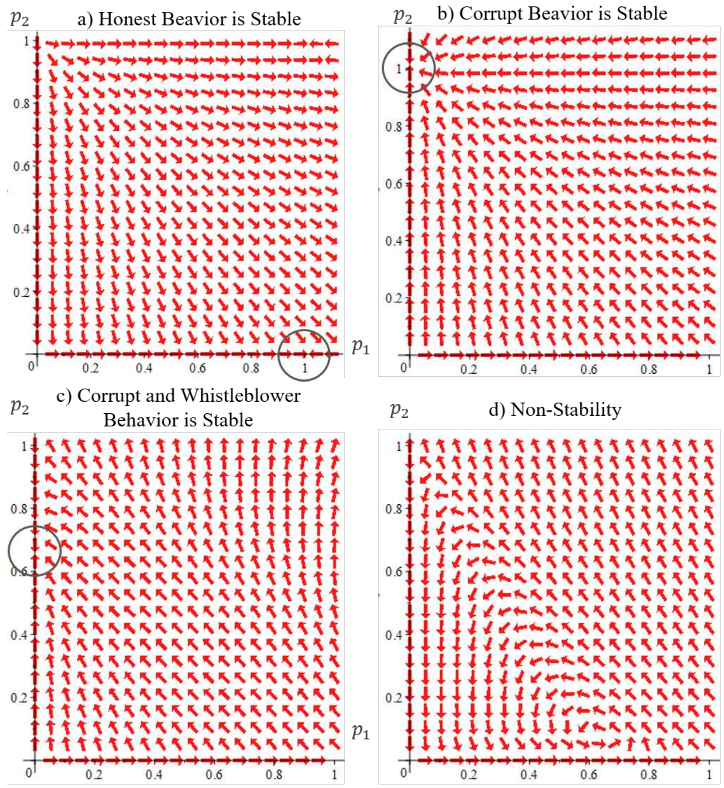
Although it is never preferable for an organization to reach the case of only corrupt employees, the case of only honest employees can be extremely expensive and even impossible to achieve, in particular when the bribe β is high or when the probability of an external monitoring agency detecting corruption, θ, is very low. In these cases, for the organization, it may be more efficient to control corruption but not eliminate it altogether. A firm with only honest employees may have salaries $w \geq w_0 - \beta - f + \frac{\beta}{\theta}$. When the firm has whistleblowers and corrupt employees, in comparison to the salaries paid when the firm only has honest employees, efficient salaries are reduced by $\Theta\frac{\sigma-\tau}{\theta}$. However, the firm has the additional costs of corruption $p_2\beta - f(\theta + p_3\Theta)$. Comparing both costs reveals that when $p_2 \leq \frac{\Theta(\sigma-\tau)+f\theta(\Theta+\theta)}{\theta(\beta+f\Theta)}$, it is more efficient to allow corrupt and whistleblower employees; that is, the higher the prize for monitoring σ is or the higher the penalty f when a corrupt employee is detected is, the more likely it is that there is slack for an increase in the proportion of corrupt employees. On the other hand, the higher the reporting costs τ is or the higher the payment for her illicit acts β is, the lower the proportion of corrupt employees the firm may wish to have.

5.5 The stability of honest, corrupt and whistleblowing employees

An environment where honest, corrupt and whistleblowing employees coexist occurs when $p_1^* = a$, $p_2^* = b$, with $a, b > 0$ and $a + b < 1$. However, this equilibrium point is not stable. The model is asymptotically stable under the circumstances described above for an environment of all honest employees, an environment of all corrupt employees or an environment in which employees are either corrupt or whistleblowers. Intuitively, this is easy to show since if the reward for whistleblowing is higher than the cost, then being honest is dominated by whistleblowing. On the contrary, if the costs of whistleblowing are greater than the reward, then whistleblowing is dominated by a strategy of being honest. Thus, in the long term, employees can be only honest, only corrupt, or corrupt and whistleblowers; honest behavior and whistleblowing can never coexist in the long run.

Figure 3 shows a summary of the phase diagrams for the following cases: a) honest behavior is stable, b) corrupt behavior is stable, c) whistleblowers and corrupt employees become stable and d) equilibrium is never achieved.

Figure 3: Phase diagram



6 Concluding remarks and policy recommendations

In this study, we have examined the role of whistleblowing in and potential impact of whistleblowing on the persistence of corruption in organizations as a mechanism to control wrongdoing. In particular, we have modeled the costs and implications of such a policy in the short run and long run. Our analysis draws some conclusions and recommendations for anticorruption policies in organizations in the context of whistleblowing behavior.

First, both static and evolutionary game approaches show that whistleblowing as a mechanism to control wrongdoing is only relevant under the existence of external monitoring. When an external enforcing agency is not in place, the organization will have corrupt employees regardless of the existence of whistleblowers. If the probability of detecting wrongdoing with an external mechanism is close to zero, then in the long term, all employees will begin to behave corruptly.

Second, whistleblowers reduce the minimum wages required to avoid corruption within an organization, making it less costly for an organization to combat corruption. This result shows that under some scenarios, whistleblowing can be a very effective and efficient way to deter corruption compared to the more traditional economic view.

Third, an increase in the whistleblowers' reward implies that the incentives for wrongdoing are less attractive, so a percentage of corrupt employees will prefer to start behaving honestly. At the same time, as there is a smaller proportion of corrupt employees, it is less attractive to behave as a whistleblower, which also decreases the proportion of whistleblowers. If bribery increases, behaving corruptly is more attractive than being a whistleblower, and a proportion of whistleblowers begin to behave corruptly, while another proportion of whistleblowers migrate toward behaving honestly. These results point to the fact that whistleblowing strategies seem to be less attractive for activities with very high bribery in comparison to the rewards for whistleblowers, for example, manufacturing or retail, but not for financial services in general.

Finally, in the long run, our main conclusion in the context of our strategic environment, is that honest, corrupt and whistleblowing employees cannot coexist, as this equilibrium point is never stable. Hence, in the long run, it is possible to envision three scenarios: i) all honest employees, ii) all corrupt employees or iii) employees being either corrupt or whistleblowers. The intuition behind this result is that if the reward for blowing the whistle is higher than the cost, then being honest is dominated by being a whistleblower. On the contrary, if the costs of blowing the whistle are greater than the reward, then whistleblowing is dominated by a strategy of being honest.

References

Acemoglu, D. (1995). Reward structures and the allocation of talent. *European Economic Review*, *39*(1), 17–33.

Acemoglu, D., & Jackson, M. O. (2017). Social norms and the enforcement of laws. *Journal of the European Economic Association*, *15*(2), 245–295.

Aidt, T. S. (2003). Economic analysis of corruption: a survey. *The Economic Journal*, *113*(491), F632–F652.

Alleyne, P., Hudaib, M., & Pike, R. (2013). Towards a conceptual model of whistle-blowing intentions among external auditors. *The British Accounting Review*, *45*(1), 10–23.

Anastasopoulos, N. P., & Anastasopoulos, M. P. (2012). The evolutionary dynamics of audit. *European Journal of Operational Research*, *216*(2), 469–476.

Andvig, J. C., & Moene, K. O. (1990). How corruption may corrupt. *Journal of Economic Behavior & Organization*, *13*(1), 63–76.

Arce, D. G. (2010). Corporate virtue: Treatment of whistle blowers and the punishment of violators. *European Journal of Political Economy*, *26*(3), 363–371.

Baker, G., Gibbons, R., & Murphy, K. J. (1994). Subjective performance measures in optimal incentive contracts. *The Quarterly Journal of Economics*, *109*(4), 1125–1156.

Becker, G. S. (1968). Crime and punishment: An economic approach. In *The economic dimensions of crime* (pp. 13–68). Springer.

Becker, G. S., & Stigler, G. J. (1974). Law enforcement, malfeasance, and compensation of enforcers. *The Journal of Legal Studies*, *3*(1), 1–18.

Block, M. K., & Heineke, J. M. (1975). A labor theoretic analysis of the criminal choice. *The American Economic Review*, *65*(3), 314–325.

Cadot, O. (1987). Corruption as a gamble. *Journal of Public Economics*, *33*(2), 223–244.

Cai, G., & Kock, N. (2009). An evolutionary game theoretic perspective on e-collaboration: The collaboration effort and media relativeness. *European Journal of Operational Research*, *194*(3), 821–833.

Cheng, J., Bai, H., & Yang, X. (2017). Ethical leadership and internal whistleblowing: A mediated moderation model. *Journal of Business Ethics*, 1–16.

Ehrlich, I. (1973). *The deterrent effect of capital punishment: A question of life and death.* National Bureau of Economic Research Cambridge, Mass., USA.

Freeman, R. B. (1996). Why do so many young american men commit crimes and what might we do about it? *Journal of Economic perspectives*, *10*(1), 25–42.

Hellmann, T., & Staudigl, M. (2014). Evolution of social networks. *European Journal of Operational Research*, *234*(3), 583–596.

Heyes, A., & Kapur, S. (2008). An economic model of whistle-blower policy. *The Journal of Law, Economics, & Organization*, *25*(1), 157–182.

Hofbauer, J., & Weibull, J. W. (1996). Evolutionary selection against dominated strategies. *Journal of economic theory*, *71*(2), 558–573.

Liu, S.-m., Liao, J.-q., & Wei, H. (2015). Authentic leadership and whistleblowing: Mediating roles of psychological safety and personal identification. *Journal of Business Ethics*, *131*(1), 107–119.

Lui, F. T. (1986). A dynamic model of corruption deterrence. *Journal of public Economics*, *31*(2), 215–236.

Murphy, K. M., Shleifer, A., & Vishny, R. W. (1991). The allocation of talent: Implications for growth. *The quarterly journal of economics*, *106*(2), 503–530.

Near, J. P., & Miceli, M. P. (2016). After the wrongdoing: What managers should know about whistleblowing. *Business Horizons*, *59*(1), 105–114.

Rose-Ackerman, S. (1975). The economics of corruption. *Journal of public economics*, *4*(2), 187–203.

Sah, R. K., & Stiglitz, J. E. (1988). Committees, hierarchies and polyarchies. *The Economic Journal*, *98*(391), 451–470.

Søreide, T. (2008). Beaten by bribery: why not blow the whistle? *Journal of Institutional and Theoretical Economics JITE*, *164*(3), 407–428.

Tirole, J. (1996). A theory of collective reputations (with applications to the persistence of corruption and to firm quality). *The Review of Economic Studies*, *63*(1), 1–22.

Villena, M., & Villena, M. (2010). On the economics of whistle-blowing behaviour: The role of incentives. *Online at https://mpra.ub.uni-muenchen.de/36310/ MPRA Paper No. 36310*.