Handout for "Disclosure, Private Revelation or Silence: Whistleblowing Incentives and Managerial Policy"

Bo Cowgill, Tarek Ghani

November 13, 2009

1 Research Questions and Results

How should managers design incentives in response to possible public revelation of organizational failure?

- High-type (socially responsible) only punish silence. Sufficiently large penalty leads all employees to report violation privately.
- Low-type (profit maximizing) punish private reporting and whistleblowing for low and moderate violations, choosing to gamble against risk of whistleblowing or natural revelation.

What sorts of wrongdoing are most likely to be reported publicly rather than privately?

- High violations are always fixed by both types and all employes always report privately.
- Moderate violations always fixed by both types, but low-type manager prefers not to be informed and take gamble. Some employees then choose to stay silent with low-type manager.
- Low violations are only fixed by high-type manager, leading to whistleblowing when ethical employee faces low-type manager.

2 Model details

Basic elements of policy:

- **Actors**: Employee, Manager, Firm and Society
- **Actions**:
  - Employees can stay silent, report privately or whistleblow.
  - Managers can ignore violation or fix.
- **Payoffs**:
  - At the end of each game, we calculate the payoffs to society and the firm.
  - The employee and managers payoffs are a function of society and the firms payoffs, weighted by their type.

– "Higher" type managers/employees place a greater weight on society's payoffs.

Sequences of events

1. Nature draws the violation v and the agents' types s and t. t and v are revealed to the employee. s is revealed to the manager.
   – Violations $v \in [0,1]$. v is the social cost of the violation. $v = 0$ means no violation and no cost imposed upon society. Violation is observed privately by the employee.
   – G the known exogenous distribution of violations with *strictly positive* probability at every $v \in (0,1)$.
   – Manager type is $s \in \{0,1\}$. Privately known. $s = 1$ is a "high" type that cares more about society, zero is the opposite.
   – Employee's type is $t \in [0,1]$. Privately known. Higher ts refer to higher concern about society.
   – β is the prior probability that the manager is type $s = 1$. It is commonly known.
   – η is the prior pdf on the employees' type t. It is assumed to be differentiable.
2. Manager announces a *whistleblowing policy* $C(h,s)$, ie, makes it public. We assume that the manager will commit to his announced whistleblowing policy.
   – $C(h,s) \geq 0$ is a whistleblowing policy specifying a schedule of penalties imposed on the employee, conditional on h, the events observable to the manager, and s, the manager's type.
   – For any given (h,s), $C(h,s)$ is bounded by $\bar{c} \geq C(h,s)$.
   – The imposition of these penalties is costless to the manager and therefore credible.
3. Each employee chooses his response $a(v,t)$ based on the violation v and the employee's type t.
   – $a_e(v,t) \in \{\phi, p, w\}$ refers to the employee's action contingent on the violation v and the employee's type t. The employee's actions comes from a set $\{\phi, p, w\}$ which can be intrepreted as follows:
     * ϕ refers to doing nothing.
     * p refers to revealing the violation v to the manager privately.
     * w refers to whistleblowing (ie revealing the violation publicly).

     Note that $a_e(v,t)$ does not depend on the manager's type s or the manager's announced whistleblowing policy $C(h,s)$.

AT THIS POINT, THE GAME SPLITS DEPENDING ON THE EMPLOYEE'S ACTION

If the employee does nothing ($a_e(v,t) = \phi$):

4. If the employee does nothing, Nature decides whether to make the violation common knowledge to society at large (including the manager).
   – This will happen with probability $q_\phi v$ with $q_\phi \in [0,1)$.
   – If nature reveals the violation, then $\Omega_\phi \in \{0,1\} = 1$.
   – The probability that the violiation is revealed *is increasing* in the size of v.
5. Payoffs are distributed.
   – In this scenario, **society's** payoffs are $-\delta v$ ($\delta > 0$) if $\Omega_\phi = 1$ and $-v$ otherwise. You can think of δ as representing the repultational harm to the firm (which has negative externalities on society as a whole).

- The **firm's** payoffs are $-(\alpha + \delta)v$ if Nature reveals the violation publicly. Otherwise it is 0.
- The **manager's** payoffs π_m are $s\pi_S + (1-s)\pi_F$. The manager's type s represents how much he weighs society vs his firm's benefit.
- The **employee's** payoffs π_e are $t\pi_S + (1-t)(\pi_F - C)$. Her type t represents how much she weighs society vs his firm's benefit. The intuition behind this is that the employee will benefit in proportion to his firm minus the punishment he gets from violating the firm's policy.

If the employee decides to reveal the violation privatley to the manager ($a_e(v,t) = p$):

4. The manager now decides an action $a_m(v,s) \in \{f, \neg f\}$ (fix or not fix) depending on the violation v and the manager's type s. Fixing a violation v costs the firm αv, where $\alpha > 0$.
5. Nature now decides whether to reveal the violation v publicly (denoted $\Omega_p \in \{1,0\}$). $Pr[\Omega_p = 1|v] = q_p v \in (q_\phi, 1)$.
   - Note: By assumption, $q_\phi < q_p$.
   - Note: *If the manager fixes the problem, the violation never becomes public.*
6. Payoffs are distributed.
   - Here, society's payoffs π_S are 0 if the manager fixes ($a_m = f$). If the manager doesn't fix and Nature reveals the violation, society's payoffs are $-\delta v$. Otherwise society's payoff is $-v$.
   - The firm's payoffs π_F are $-\alpha v$ if the manager fixes the violation. If the manager doesn't fix the violation and Nature reveals the outcome to the public, then the firm's payoffs are $-(\alpha+\delta)v$ (note both a reputational and pecuinary cost). Otherwise, the firm's payoffs are 0.
   - As in the case above, the manager's payoffs π_m are $s\pi_S + (1-s)\pi_F$.
   - As in the case above, the employee's payoffs π_e are $t\pi_S + (1-t)(\pi_F - C)$.

If the employee whistleblows ($a_e(v,t) = w$):

4. v becomes common knowledge ($\Omega_w = 1$)
5. The management fixes the problem by paying αv directly (and are later harmed reputationally by δv and punishes the employee with C.
6. Payoffs are distributed.
   - Society's payoffs are $-\delta v$.
   - The firm's payoffs are $-(\alpha + \delta)v$.
   - As in the case above, the manager's payoffs π_m are $s\pi_S + (1-s)\pi_F$.
   - As in the case above, the employee's payoffs π_e are $t\pi_S + (1-t)(\pi_F - C)$.

3 Revisiting Main Conclusions

1. Violations can be partitioned into three sets, of which only the "lowest" need be non-empty;
2. The "highest" violations are invariably reported privately by all employee types to either type of manager who surely fixes the violation and does not penalize the employee;

3. "Moderate" violations are fixed by both types of manager conditional on being privately informed, but only the high type manager prefers to be told of such violations whereas the low type manager penalizes any private reporting of "moderate" or "low" violations;

4. Only the high type manager fixes "low" violations when informed but not all employee types are willing to report all such violations; the low type manager prefers not to be informed and does not fix the violation if the employee nevertheless reports it privately; and sufficiently high employee types blow the whistle on "low" violations;

5. Only the low type manager penalizes whistleblowing or private reporting; the high type manager only penalizes remaining silent. Moreover, not all low or moderate violations are reported to either type of manager, even when there are no penalties for any action.