EC951 Lecture 3: Moral Hazard and Teams

Largely based on "Moral Hazard in Teams", Holmstrom, Bell Journal of Economics, 1982.

David Reinstein

October 30, 2011

Holstrom (1982)

Main points (note there are 2 related parts to the paper):

- With uncertainty and risk aversion, inducing optimal effort requires 'monitoring' individual agents' performance, and payments based on a 'sufficient statistic.' But this monitoring can be expensive, and making rewards individual might deter agents from fruitful cooperation.
  - Principal can compare or 'benchmark' agents' performance to one another – this can induce optimal effort when all uncertainties are the same among agents (e.g., a good/bad market, weather, etc).
  - With a common shock, observing different outputs for different agents means that one agent put in more effort!
  - But even with such benchmarking, principal will have to pay a risk premium for 'idiosynchratic' risk (this contrasts with asset market pricing); pooling risk among agents could lessen this.

- On the other hand, when only the team's output is observed and rewarded, there is a potential incentive to 'free ride'.
  - This 'public goods problem' can be 'solved' if we can implement a contract that involves an injection or extraction of surplus (reward/punishment larger than actual profit or loss)
  - This can only be accomplished with an external principal

Free rider problem : intuition

Say n of us are collaborating to produce research reports.
The boss pays our team based on the number of good reports we produce, for which we bill the client $\$1$ per report.

She says 'you all will get the same amount, $\frac{\$1}{n}$ for each report, and I want you all to put in the same effort.' She doesn't want to get involved in disputes: we have to work it out between ourselves.

My effort (e_i) is costly, increasingly costly as I work additional hours – say it costs me $v_i(e_i)$. Assume (WLOG) that my productivity remains constant at x per hour.

If I were producing these by myself I'd work until the marginal effort required to produce a report equalled my pay for the last report, and then stop. I.e., until marginal cost equals marginal benefit, $v_i'(e_i) = 1$. This is technically efficient.

Let's say the resulting output is $q^* = 1$ per person who works efficiently.

But, in the team, with the incentives above, I only get $\frac{1}{n}$ of the pay from any report I write. So I will only work until $v_i'(e_i) = \frac{1}{n}$. I will 'slack off'.

This is the classic ‘free rider’ problem in public goods models. Because of this problem, we the employees are worse off.

This is not Pareto-efficient: if any of us could work and keep all of the gains he could make himself better off and all the others would be the same.

Mechanism for free rider problem: intuition

Unless we can observe each individual's effort and reward/punish him based on it (which may be difficult to enforce), as a team we can not solve this problem between ourselves.

If everyone is to be paid the same, we need to increase every team member's incentive by a factor of n so that $v_i'(e_i) = n\frac{1}{n} = 1$.

But there is no way to do this: if we *do* perform efficiently we produce $nq^* = n$ units and are paid $\$nq^* = \$n$ in total.

But to increase incentives enough we'd have to pay everyone $\$n\frac{1}{n} = \$1$ per unit of our total production, which would cost us $\$n * n$ – where would we get the money to pay ourselves this?

Collective reward and punishment

If we could commit to us all having to throw away all of our pay if we produce less than n total reports we would also have this efficient incentive.

As long as all others were putting in optimal effort it would be your best response to do so as well.

But let's say one person slacked off (was lazy) and we still received some pay: committing to throw away this money may not be feasible!

However, the boss could implement these incentives! She could pay us nothing if we fail to produce the n total reports! That would get our butts in gear (get us to work).

Model

n agents, denoted by $i = 1, 2...n$. Each agent i chooses an effort level e_i.
Individual efforts cost $v_i(e_i)$, where $v_i'(.) > 0$ and $v_i''(.) > 0$.
Profile of efforts $\mathbf{e} = (e_1, e_2, ..., e_n) \in A$ generates group output, a function: $x : \mathbf{e} \rightarrow R$.

Principal observes group output but not individual effort.

Let $s_i(x) \geq 0$ be the payment to agent i as a function of output x

Budget balance: for all x, $\sum_i^n s_i(x) = x$.
Social welfare: $x(\mathbf{e}) - \sum_{i=1}^n v_i(e_i)$.
Socially efficient effort level profile $\mathbf{e}^*$ must satisfy

$$\mathbf{e}^* \in \text{argmax}_{\mathbf{e} \in A}[x(\mathbf{e}) - \sum_{i=1}^{n} v_i(e_i)] \quad (1)$$

Is there a balanced budget scheme that attains efficiency?

Theorem

Holmstrom (1982): There does not exist a continuously differential payment scheme, s, which satisfies budget balance and yields an efficient level of effort in Nash equilibrium among agents.

Proof: In an interior Nash equilibrium, for any agent the first order condition must satisfy $\frac{\partial s_i(x)}{\partial x}\frac{\partial x(e)}{\partial e_i} - v_i'(e_i) = 0$

However, an interior Pareto optimal effort level, $\mathbf{e}$, satisfies $\frac{\partial x(\mathbf{e})}{\partial e_i} - v_i'(e_i) = 0 \; \forall i$

Thus, for the Nash equilibrium to be also efficient it must be the case that $\frac{\partial s_i(x)}{\partial x} = 1$ for every i, implying $\sum_{i=1}^{n} \frac{\partial s_i(x)}{\partial x} = n$.

However, budget balance requires $\sum_i^n s_i(x) = x$. Differentiating both sides of this yields $\sum_{i=1}^{n} \frac{\partial s_i(x)}{\partial x} = 1$ for all x.

This contradicts the requirement (for an efficient Nash equilibrium) that $\frac{\partial s_i(x)}{\partial x} = 1$ for all i. Q.E.D.

Relaxing the budget balance requirement

Def: a payment scheme is *feasible* if $s_i(x) \geq 0$ for all x and for all i, and $\sum_i^n s_i(x) \leq x$, for all x.

Theorem

There exists a feasible payment scheme $s^*(x)$ such that the efficient effort level is a Nash equilibrium among agents.

Proof.

Let efficient effort level be given by $\mathbf{e}^*$. Define the payment scheme as follows: $s_i(x) = b_i$ if $x \geq x(\mathbf{e}^*)$, and $s_i(x) = 0$ otherwise. Let $\sum_i^n b_i = x(\mathbf{e}^*)$. Let $b_i \geq v_i(e_i^*) > 0 \; \forall i$.This can be achieved since $\mathbf{e}^*$ is the efficient effort level.

Lemma

In the payment scheme above, choosing efficient actions is a Nash equilibrium:

Proof of lemma: Individual payoff from doing efficient effort, e_i^*, given that everyone else does so, is $u_i(s_i^*, e_i^*, e_{-i}^*) = s_i(\mathbf{e}^*) - v_i(e_i^*) = b_i - v_i(e_i^*) > 0$.

Payoff from deviating and choosing a lower effort e_i is given by $0 - v_i(e_i) \leq 0$. Thus shirking is not profitable.

Working harder than e_i^* is not worthwhile since it has no bearing on reward, which remains at level b_i, while the cost of effort increases. □

Caveat: everyone putting in zero effort is also a Nash equilibrium under this scheme.

The above set of results show that the ability to commit to throw away output (or have a third-party 'principal' capture it) in some states of the world is very useful to implement socially efficient effort levels. This in turn suggests that some forms of organization such as labour managed firms or partnerships may find it harder to attain efficient outcomes as compared to capitalist firms.

Practice question: 1.4 in “moral hazard practice questions.pdf”