# Module 8: Multi-Agent Models of Moral Hazard

Information Economics (Ec 515) · George Georgiadis

**Types of models:**

1. No relation among agents.
   - Can many agents make contracting easier?
2. Agents' shocks are correlated.
   - *e.g.*, output of agent $i$ is given by $q_i = a_i + \epsilon_i$ and $\epsilon_i$'s are positively correlated.
   - Output of agent $i$ "contains" information about the efforts of all the other agents.
3. Joint production.
   - Try to separate the agents' performance given joint output.
4. Each agent may have information about the effort choices of the other agents.

## Problem Formulation (with 2 agents):

- 2 risk-neutral agents
- Agent $i$ takes action $a_i$ at cost $c(a_i)$
- Agent $i$'s utility: $u_i = w_i - c(a_i)$
- Output of agent $i$: $q_i = a_i + \epsilon_i$, where $\epsilon_1$ is independent of $\epsilon_2$, and $\mathbb{E}[\epsilon_i] = 0$.

- First best:

$$\max_{a_i, w(\cdot)} \mathbb{E}[q - w(q) \mid a_i]$$
$$\text{s.t. } \mathbb{E}[w(q) - c(a_i) \mid a_i] \geq \bar{u} \quad \text{(IR)}$$

- Straightforward that (IR) binds. Then $\mathbb{E}[w(q)] = c(a_i) + \bar{u}$.
- Then the maximization problem reduces to:

$$\max_{a_i} \{\mathbb{E}[q - c(a_i) \mid a_i] - \bar{u}\} = \max_{a_i} \{a_i - c(a_i) - \bar{u}\}$$

  - FOC: $c'(a_i) = 1$.

**Implement with a Tournament**

- Consider a two-player tournament where the winner is the player who produces the highest $q$.
- Prizes $w_H$ and $w_L$ for the agent with the higher and lower output, respectively.
- *Idea:* Contestants pre-commit their investments early in life, knowing the prizes and the rules of the game.

- $\Pr\{\text{i wins}\} = \Pr\{q_i > q_j\} = \Pr\{a_i + \epsilon_i > a_j + \epsilon_j\} = \Pr\{\epsilon_j - \epsilon_i < a_i - a_j\} = H(a_i - a_j)$, where $H(\cdot)$ is the cdf of $\epsilon_j - \epsilon_i$.

- Agent $i$'s problem:

$$\max_{a_i} \; w_H H(a_i - a_j) + w_L [1 - H(a_i - a_j)] - c(a_i)$$

- FOC: $(w_H - w_L) h(a_i - a_j) = c'(a_i)$
- Symmetric equilibrium $\Longrightarrow a_i = a_j = a^* \Longrightarrow c'(a^*) = (w_H - w_L) h(0)$.
  - Each agent "wins" with probability $\frac{1}{2}$.

(1) Equilibrium effort is first best if $(w_H - w_L) h(0) = 1$.

(2) For (IR) to be satisfied, $w_H$, $w_L$ must satisfy: $\frac{w_H + w_L}{2} - c(a^*) = \bar{u}$.

- Obtain $w_H$, $w_L$ by solving (1) and (2).

- *Advantages:*

- Simplicity
- Private evaluation robustness
  - With individual incentives, if explicit contracts cannot be written down, firms may not pay the bonus. (May be solved by repeated interaction, but only partially.)
  - If a prize must be given, might as well give it to the "best" performer.

- *Disadvantages:*
  - Asymmetric equilibria
  - Risk aversion causes problems
  - Solution depends on the specification of $h(\cdot)$
  - Collusion
  - Sabotage
  - If uncertainty unfolds gradually and agents can change their efforts (after observing the realization of shocks), effort decreases as the gap between the leader and the laggard increases: Tournaments can undermine incentives in a dynamic setting.

**Holmstrom (Bell Journal, 1982)**

- $n$ risk-neutral agents, each with reservation utility $\bar{u}$.
- Agent $i$'s utility: $u_i = t_i - c(a_i)$
- Action $a_i \in A_i \subseteq \mathbb{R}$, and $c(\cdot)$ is convex.
- Output (deterministic): $q(\mathbf{a})$, where $\mathbf{a} = \{a_1, .., a_n\}$, $q(\cdot)$ is differentiable and $\frac{dq}{da_i} > 0$.
  - *e.g.*, $q = \sum_{i=1}^{n} a_i$

**First best:**

- $a^* \in \arg\max \{q(\mathbf{a}) - \sum_i c(a_i)\}$
  - Any internal maximum satisfies $\frac{dq(\mathbf{a}^*)}{da_i} = c'(a_i^*) \quad$ for all $i$.
  - Assume $q(\mathbf{a}^*) - \sum_i c(a_i^*) \geq \sum_i \bar{u}_i$.
  - Split proceeds $\{t_i^*\}_i$ such that (i) $t_i^* - c(a_i^*) \geq \bar{u}_i$ for all $i$ and (ii) $\sum_i t_i^* = q(\mathbf{a}^*)$.

**Moral Hazard Problem:**

- Use output sharing rule $\{t_i(q)\}_i$ such that $\sum_i t_i(q) = q$ (balanced budget).
  - Assume that $t_i(q)$ is differentiable.
- Agent $i$'s Problem:

$$\max_{a_i} \; t_i(q(a_i, \tilde{a}_{-i})) - c(a_i)$$

  - FOC: $t_i'(q(\tilde{\mathbf{a}})) \frac{dq(\tilde{\mathbf{a}})}{da_i} = c'(\tilde{a}_i)$
  - $t_i'(q)$ is agent $i$'s marginal pay per-unit of output.
- Can we implement first best $a^*$?
  - From the first best FOC and each agent's FOC, it must be the case that $t_i'(q(\mathbf{a}^*)) = 1$ for all $i \Longrightarrow t_i(q) = q - F_i$.
  - But then, the budget balance constraint is violated; *i.e.*, $\sum_i t_i(q) = nq - \sum_i F_i = q$ cannot hold for all $q$.
  - *To obtain first-best, every agent must get his marginal \$, but this is impossible*!)

  $\Longrightarrow$ there exists no budget balanced sharing rule that achieves first best.
- Intuition: Each agent must share the marginal benefit of his output, but he alone bears its cost.

**How to obtain First Best ?**

1. Destroy output:
   - Let:

$$t_i(q) = \begin{cases} t_i^* & \text{if } q = q(\mathbf{a}^*) \\ -K & \text{otherwise .} \end{cases}$$

   - Problems:
     (a) Not "renegotiation proof".
     (b) What if output is random? (Multiple equilibria.)
2. Budget breaker:
   - Introduce $(n+1)^{th}$ agent.

- Let:

$$t_i(q) = q - F_i \quad \text{for all } i \in \{1,..,n\}$$
$$t_{n+1}(q) = q - \sum_{i=1}^{n} t_i(q) = \sum_i F_i - (n-1)q$$

where $F_i$ are transfers from agent $i$ to the $(n+1)^{th}$ agent.

- Choose $\{F_i\}_i$ such that $t_{n+1}(q(\mathbf{a}^*)) = 0$; *i.e.*, $\sum_i F_i = (n-1)q(\mathbf{a}^*)$.
- Problems:
  (a) How to interpret budget breaker? (Not a manager. Observe that BB pays more, the lower the output.)
  (b) BB has incentives to sabotage.

3. Spotting Individual Deviations:

- Suppose $A_i$ is discrete: $q(\mathbf{a}) \neq q(\mathbf{a}')$ for all $\mathbf{a} \neq \mathbf{a}'$.
- Use the following scheme:

$$t_i(q) = \begin{cases} t_i^* & \text{if } q = q(\mathbf{a}^*) \\ -K & \text{if } q = q\left(a_i, \mathbf{a}_{-i}^*\right) \neq q(\mathbf{a}^*) \\ \frac{q+K}{n-1} & \text{if } q = q\left(a_i^*, \mathbf{a}_{-i}\right) \neq q(\mathbf{a}^*) \\ \frac{q}{n} & \text{otherwise .} \end{cases}$$

**Different Types of Implementation**

- 2 agents
- Effort $a \in \{L, H\}$; cost of effort $c_L = 0$ and $c_H = C > 0$.
- Project succeeds or fails and $\Pr\{\text{success}\} = P(x)$, where $x = \#$ of agents who exert $a = H$.
  - $P(x)$ increases in $x$.
  - Increasing returns: $P(2) - P(1) > P(1) - P(0)$ (*i.e.*, agents' efforts are complements).
- What is the cheapest way for a principal to incentivize workers?
  - Contract for worker $i$: $w_i \mathbf{1}_{\{\text{success}\}}$ (agent is protected by limited liability)

1. **Partial Implementation:**

   - Choose $w_i$'s such that there exists an equilibrium in which both agents work.
   - Assume agent $i$ believes that agent $-i$ will work. Then

$$(\text{IC}_i) \quad w_i P(2) - C \geq w_i P(1) \Longrightarrow w_i \geq \frac{C}{P(2) - P(1)} = w^P \tag{1}$$

     - The other agent faces the same constraint.
   - Suppose that each agent receives $w^P$ when the project succeeds.
   - What happens to contract given in (1) if agent $i$ beliefs that agent $-i$ will shirk?
     - Write agent $i$'s IC:

$$\begin{aligned} w^P P(1) - C &\geq w^P P(0) \\ \Longrightarrow \frac{P(1)}{P(2) - P(1)} C - C &\geq \frac{P(0)}{P(2) - P(1)} C \\ \Longrightarrow P(1) - P(2) + P(1) &\geq P(0) \\ \Longrightarrow P(2) - P(1) &\leq P(1) - P(0) \end{aligned}$$

     - Contradicts the assumption that efforts are complements
     - Therefore, if an agent believes that the other agent will shirk, then he will also shirk (*i.e.*, 2 Nash equilibria).

2. **Full Implementation:**

   - How can we ensure that both agents exerting $a = H$ is the unique equilibrium?
     - One possibility is to set $w_1 = w_2 = \frac{C}{P(1) - P(0)}$. Can we do better?
   - Yes!
     - Choose $w_1$ such that agent 1 finds it optimal to exert $a = H$ no matter what.
     - Then set $w_2 = w^P = \frac{C}{P(2) - P(1)}$. Given that agent 1 exerts $a = H$, agent 2 will also exert $a = H$.
   - To ensure that agent 1 exerts $a = H$ no matter what, we need:

$$\begin{aligned} w_1 P(1) - C &\geq w_1 P(0) \\ \Longrightarrow w_1 &\geq \frac{C}{P(1) - P(0)} \end{aligned}$$

and $w_1 \geq w^P$. Because $\frac{C}{P(1)-P(0)} > w^P$, we set $w_1^F = \frac{C}{P(1)-P(0)}$ and $w_2^F = \frac{C}{P(2)-P(1)}$.

- *Full implementation:* Concerned with characterizing all Nash equilibria.
- *Partial implementation:* Characterizing one (of possibly many) Nash equilibria.

3. **Sequential Implementation:**

- Suppose agent 1 chooses $a_1$.
- Agent 2 observes agent 1's choice and chooses $a_2$. (We assume that effort is observable but not contractible.)
- Working backwards:

$$w_2^S = \frac{C}{P(2) - P(1)},$$

*i.e.*, agent 2 finds it optimal to work when agent 1 works.

- We want to choose the wage of agent 1 such that he works, and as a consequence, agent 2 also works.
  - Agent 1's IC constraint: $w_1^S P(2) - C \geq w_1^S P(0) \Longrightarrow w_1^S = \frac{C}{P(2)-P(0)}$.
- Because $P(1) > P(0)$, $w_1^S > w_2^S$.
- There exists a unique equilibrium in which both agents work.

- Observe that $2w^P < w_1^S + w_2^S < w_1^F + w_2^F$.

# References


Board S., (2011), Lecture Notes.

Bolton and Dewatripont, (2005), *Contract Theory*, MIT Press.

Holmstrom B., (1982), "Moral Hazard in Teams", *Bell Journal of Economics.*

Winter E., (2004), "Incentives and Discrimination", *American Economic Review.*

Winter E., (2006), "Optimal Incentives for Sequential Production Processes", *RAND Journal of Economics.*