EconS 503 - Moral Hazard in Teams

In this section we analyze a standard free-rider problem of team production. First, we show that when each worker's effort is unobservable but the benefits from the project are equally distributed among all agents, equilibrium effort is suboptimal relative to the social optimal (first best), as workers free ride on each others' effort. Secondly, we analyze if the presence of a principal can help the team reach a first best outcome.

Model

Consider n agents working in a partnership to produce aggregate output

$$Q = Q(a_1, a_2, \cdots, a_n),$$

where $a_i \in [0, \infty)$ is the agent i's action, e.g., his effort. Assume that function $Q(\cdot)$ satisfies

$$\frac{\partial Q}{\partial a_i} > 0,\ \frac{\partial^2 Q}{\partial a_i^2} < 0,\ \text{and } \frac{\partial^2 Q}{\partial a_i \partial a_j} \geq 0$$

That is, every agent i's effort is productive, but at a decreasing rate. In addition, the marginal contribution of each agent's effort to the project increases in other agent's efforts, i.e., efforts are strategic complements of each other.

Suppose that each agent is risk-neutral: with wage w_i and effort a_i, the agent i's utility is

$$w_i - g_i(a_i)$$

with $g_i(\cdot)$ strictly increasing and convex in effort. Since each agent's effort or individual output is not observable, a contract in the partnership can only depend on the aggregate output:

$$w(Q) = (w_1(Q), w_2(Q), \cdots, w_n(Q))$$

We require each wage function $w_i(\cdot)$ to be differentiable (almost everywhere). All agents in the partnership share the aggregate output in the following sense:

$$\sum_{i=1}^{n} w_i(Q) = Q \text{ for each } Q$$

A key observation in this model is the positive externality among agents: if all agents are rewarded when Q increases, then one agent working hard will increase the reward to all other $n-1$ agents as well. This provides a possible source for free-riding incentives, and an inefficient outcome, as we next show.

Free-riding incentives in the absence of a principal

The first-best profile of actions $\hat{a} = (\hat{a}_1, \cdots, \hat{a}_n)$ has a social planner considering aggregate welfare, and solving the following problem

$$\max_{a_1,\cdots,a_n} \sum_{i=1}^{n} (w_i(Q) - g_i(a_i))$$

Taking FOC with respect to each effort a_i yields

$$\sum_{i=1}^{n} \frac{\partial w_i(Q(\hat{a}))}{\partial Q} \cdot \frac{\partial Q(\hat{a})}{\partial a_i} - g_i'(a_i) = 0$$

or, rearranging

$$\sum_{i=1}^{n} \frac{\partial w_i(Q(\hat{a}))}{\partial Q} \cdot \frac{\partial Q(\hat{a})}{\partial a_i} = g_i'(a_i)$$

That is, at the social optimum, every agent i increases his effort a_i until his marginal disutility, $g_i'(a_i)$, coincides with the marginal benefit that his effort entails for the entire group, $\sum_{i=1}^{n} \frac{\partial w_i(Q(\hat{a}))}{\partial Q} \cdot \frac{\partial Q(\hat{a})}{\partial a_i}$.

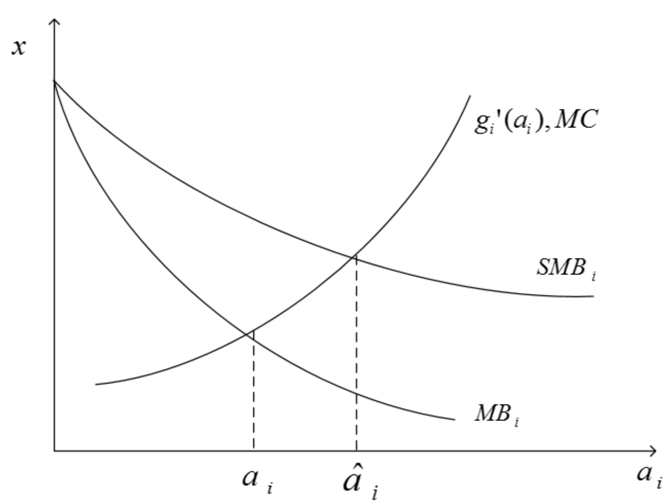
We next show that when actions are not contractible, the first-best is no longer achievable. In particular, every agent takes other agents' actions $a_{-i} = (a_1, \cdots, a_{i-1}, a_{i+1}, \cdots, a_n)$ as given, and solves

$$\max_{a_i} \; w_i(Q) - g_i(a_i)$$

Taking FOC with respect to agent i's effort, and applying the chain rule, we obtain

$$\frac{dw_i(Q(a_i, a_{-i}))}{\partial a_i} \cdot \frac{\partial Q(a_i, a_{-i})}{\partial a_i} = g_i'(a_i) \text{ for each } i$$

In words, in equilibrium agent i increases his effort level a_i until the point in which his marginal disutility, $g_i'(a_i)$, coincides with his private marginal benefit, $\frac{dw_i(Q(a_i,a_{-i}))}{\partial a_i} \cdot \frac{\partial Q(a_i,a_{-i})}{\partial a_i}$. As a consequence, he selects an inefficient effort relative to the social optimum, as depicted in the next figure, where $a_i < \hat{a}_i$.



Restoring first-best outcomes by introducing a principal

Consider now that we relax the budget-balance condition

$$\sum_{i=1}^{n} w_i(Q) = Q$$

so that a payment $B > 0$ goes to a principal that the team hired, implying

$$\sum_{i=1}^{n} w_i(Q) < Q$$

In particular, consider an incentive scheme that pays every agent i a reward $w_i(Q) = \bar{w}_i$ if total output coincides or exceeds the first-best level, $Q \geq Q_{FB}$, but zero otherwise, where $\sum_{i=1}^{n} \bar{w}_i = Q_{FB}$, and $\bar{w}_i > g(a_i^{FB})$ for every agent i. Importantly, the principal does not observe the agents' efforts levels, but only the aggregate output Q, and whether it exceeds the first best level, Q_{FB}.

Such incentive scheme is a NE. To see why, consider agent i. Taking the actions of all other agents as given, $a_{-i} = a_{-i}^{FB}$, we can now check if agent i also prefers to choose $a_i = a_i^{FB}$ or instead deviate. If agent i deviates upwards, $a_i > a_i^{FB}$, his cost of effort increases but his reward doesn't (his reward is still $\bar{w}_i$), so he does not have incentives to deviate. If he deviates downwards, $a_i < a_i^{FB}$, he saves the cost of effort but his reward drops to zero. In particular, his utility from $a_i = a_i^{FB}$ is $\bar{w}_i - g_i(a_i^{FB})$, while by deviating to $a_i < a_i^{FB}$ his utility becomes $0 - g_i(a_i)$. Hence, agent i does not want to deviate if

$$\bar{w}_i - g_i(a_i^{FB}) > 0 - g_i(a_i)$$

or,

$$\bar{w}_i > g_i(a_i^{FB}) - g_i(a_i)$$

His optimal deviation would thus be $a_i = 0$. That is, if slacking entails a zero reward, then his optimal deviation is to slack completely, which entails a zero cost of effort, $g(0) = 0$. In such a case, the above inequality becomes

$$\bar{w}_i > g_i(a_i^{FB})$$

which holds by definition. Hence, agent i does not have incentive to deviate from the first-best effort level $a_i = a_i^{FB}$.