Incentive Reversal

Eyal Winter
The Federmann Center For the Study of Rationality
NTU April 2024

Rewards and Incentives

- In a team environment rewards may affect performance in a non-monotonic way.
- The argument is not based on any behavioral or psychological assumption.
  - Gneezy Rustichini (2000)
  - Falk and Fehr (2001)
- The argument builds on the externalities among peers.

- We show that increasing rewards for all agents can result in the shrinking of the set of agents who exert effort.
- The effect can be dramatic:
  - Under the low set of incentives everyone exerts effort and under the high set of incentives only one player does so.
- The Basic Idea: High rewards can create an incentive for some players to exert effort as a dominant strategy thus may cause other to free ride.

Two Agents

  - 2-agent organization. Each agent deals with a single task.
- c is the cost of effort.
  - Effort increases the success probability of a task from α to 1.
- The project succeeds iff all tasks are successful.
  - The principal can observe only the outcome of the project.
- A **mechanism** is a pair of payoffs (v_1, v_2) that will be paid to the agents if the project succeeds.
  - If the project fails they get zero.

Example

- Consider the 2-agent benchmark model with $a = 0.9$ and $c = 1$.
- Assume agents move sequentially.
  - Assume $v_1 = 5.5$ and $v_2 = 11$.
- Under this mechanism both agents exert effort in the unique equilibrium.
  - Suppose now that the principal raises the rewards of both agents by 15%: $v^*_1 = 6.33$ and $v^*_2 = 12.66$.
- It is now a dominant strategy for agent 2 to exert effort.

Example

  - $12.66*0.9^2 < 12.66*0.9 – 1.$
- But now the first agent, who realizes that the second will invest no matter, looses his incentives:
  - $(6.33)(0.9) > 6.33 - 1.$
- So the unique equ. now is with only player 2 investing.
- So the principal paid more and got less effort.

If Agents Move Simultaneously

	c	nc	
c	5.5-1=4.5 11-1=10	5.5 *.9-1=3.95 11 * .9=9.9	Low Rewards
nc	5.5 *.9=4.95 11 * .9-1=8.9	5.5 $*0.9^2 = 4.45$ EQ $11*0.9^2 = 8.91$	

	c	nc	
c	6.33-1=5.33 12.66-1=11.66	6.33 *.9-1=4.697 12.66* .9=11.394	High Rewards
nc	6.33*.9=5.7 EQ 12.66 * .9-1=10.4	$6.33*0.9^2 =5.13$ $12.66*0.9^2=10.25$	

The Model

- N is the set of agents
- Each agent makes an effort decision
- $d_i \in \{0,1\}$ (1 means effort 0 means shirking)
- p is a function from $\{0,1\}^N$ to $[0,1]$.
- Symmetry: p:$\{0,1, \ldots , n\}\rightarrow[0,1]$, with $p(k)$ being the probability of success if k agents contribute.
- p is strictly increasing.
  - In the benchmark model $p(k) = \alpha^{n\text{-}k}$.

Mechanism

- A mechanism is a vector $v = (v_1,\ldots,v_n)$
- agent i receive the payoff v_i if the project succeeds and zero otherwise.
- We assume that agents act sequentially
- When player i decides about his effort, he observes the effort decisions of all his predecessors.
- Denote by $\mathrm{E}(v)$ the set of agents who exert effort in the SPE of the underlying game.

Incentive Reversal

- We say that p is susceptible to incentive reversal if
- there exist two reward vectors v^1 and v^2 such that $v^1 < v^2$ (in each coordinate) and nevertheless $\mathrm{E}(v^2) \subset \mathrm{E}(v^1)$

Theorem (2-agent case)

- p is susceptible to reverse incentives if and only if it involves complementary tasks; i.e,

$$p(\{i,j\}) - p(\{i\}) > p(\{j\}) - p(0)$$

- Proposition 1: If p is susceptible to incentive reversal then it involves complementary tasks.
- Proposition 2: Any technology p, which involves complementary tasks is susceptible to incentive reversal.

Proposition 1

- Suppose $v^1 < v^2$ and nevertheless $\mathrm{E}(v^2) \subset \mathrm{E}(v^1)$
- Player 2's strategies:

a. Always exert effort (dom 1)
b. Exert effort iff 1 does so (mimic)
c. Exert effort iff 1 shirks (flip)
d. Always shirk (dom 0)

- Case 1: $(1,1) = \mathrm{E}(v^1)$, so under v^1 the best response to effort by agent 1 is effort by agent 2.
- But clearly this must hold also for v^2
  - Hence $(1,0) = \mathrm{E}(v^2)$ cannot be the case
- Case 2: $(0,0) = \mathrm{E}(v^2)$, 2's strategy either dom 0 or mimic
  - But dom 0 is impossible since $(1,1) = \mathrm{E}(v^1)$
    - (Otherwise 2 will choose 0 also in $v^1 < v^2$)
  - But also mimic is impossible.
    - (Otherwise 1 would have chosen 1, as he does under v^1)

- Hence $(0,1) = \mathrm{E}(v^2)$ is the only possible case.
- So $(1,1) = \mathrm{E}(v^1)$ and $(0,1) = \mathrm{E}(v^2)$
  - The action of player 1 in the two schemes implies that the strategy of agent 2 under v^1 is to exert effort iff 1 does so.
  - Otherwise (0,1) must be an equilibrium also under v^1

- The Incentive Constraints:

$$p(2){v_2}^1 - c > p(1){v_2}^1 \text{ and } p(1){v_2}^1 - c < p(0){v_2}^1$$

- Hence,

$$c/[p(2)\text{-}\ p(1)] < {v_2}^1 < c/[p(1)\text{-}\ p(0)]$$

- Or,

$$p(2)\text{-}\ p(1) > p(1)\text{-}\ p(0)\ \text{-}\ \textit{IRS}$$

Other Options

v^1 v^2

(1,0) (0,0) impossible since in v2 player 1 would choose 1

(0,1) (0,0) impossible since in v2 player 2 would choose 1

(0,0) not possible

The General Case

- For the two agent case the technology can be either convex or concave (either supermodular or submodular)
- There are thus two candidates for the general case.

Theorem 1

- A technology p is immune to incentive reversal if and only if it has decreasing returns to scale.
- So it is enough to have some region of convexity in p for it to be susceptible to incentive reversal.

Proposition 4

- If p has increasing returns to scale, then there exist two reward vectors v^1 and v^2 with $v^2 > v^1$ such that under v^1 all agents exert effort in equilibrium, while under v^2 only one agent does so.

Implications

- A fund-raiser who elicits donations for a cause should be cautious in his campaign.
  - Suppose that donors are approached sequentially and that the cause requires a certain threshold of funds (making the fund-raising technology satisfy complementarity).
- Boosting the attractiveness of the cause in a way that would make it a dominant strategy for late movers to donate may make early movers reluctant to chip in their contributions.

- There is considerable empirical and experimental evidence on psychological peer effects showing that workers are typically reluctant to exert effort when they observe their peers shirking.
- This reluctance may in fact be quite effective in sustaining a high level of effort within teams, because it serves as an implicit threat against shirking.
- In such teams an increase in rewards may quash this implicit threat. Some agents may find it attractive enough to exert effort even when observing their peers shirking, which in turn may encourage these peers to shirk.

Experimental Results