Lecture 2: Extrinsic vs intrinsic motivation

Till now: Positive intrinsic motivation and extrinsic motivation due to incentives work in the same direction

Psychology/sociology: Incentives might be harm intrinsic motivation

"Forbidden fruit effect": Punishing an action if detected makes the action more attractive

Tom Sawyer sells permissions to paint the fence.

Natural experiment (Gneezy and Rusticchini):

Problem: Parents pick up their children at the daycare too late

"Solution" at an Israeli daycare: Small monetary fine for coming late

Result: Even more parents pick up the children too late

Lab experiment (Falk and Kosfeld):

Basic question: Is it profitable for the principal to (partially) control agent's effort choice

Payment functions:

$$
\begin{aligned}
\pi_P &= 2e \\
\pi_A &= 120 - e
\end{aligned}
$$

2 stages:
Principal chooses whether to control for and prevent low effort, i.e. $E = [\underline{e}, 120]$, or not, i.e. $E \in [0, 120]$
Agent chooses $e \in E$

Three treatments: $\underline{e} = 5$, $\underline{e} = 10$, and $\underline{e} = 20$

Results:

Hidden cost of control: For all three treatments, it holds for any value of $e > \underline{e}$ that there are always strictly more agents who choose at least e if the principal controls than if he does not.

For the principal these control costs more than outweigh the control gain of preventing $e < \underline{e}$.

The majority of the principals chooses not to control the agent.

Extension: Gift exchange game where principal sets wage w

positive $e - w$ relation

$e - w$ relation flatter when agent is controled

Explanations for crowding out of intrinsic motivation:

Incentives carry information from principal to agent (Benabou and Tirole 2003):

Basic mechanism: The principal has superior information about whow suited the agent is for the task (how difficult it is, how much he likes it, etc). Strongly powered incentives adversely impact on the agent's (worker's, child's) perception of the task, or of his own abilities, etc.

Results:
Less powered incentives are optimal
But: In equilibrium, incentives used - incomplete crowding out. Otherwise incentives would not be signal for difficulty - contradiction

Social Reputation (Benabou-Tirole 2006):

Three types of motivations:

selfishness
social preferences - altruism
reputation: social reward for being regarded as altruist, shame for being
regarded as very selfish (also self-perception)

Amount of selfishness and altruism not known by others, has to be
deduced from actual action $\Rightarrow$ The higher the incentives, the less reveals
the actual action about altruism $\Rightarrow$ Crowding out of reputational motive
by incentives

Agent chooses $a \geq 0$ contributing to a public good, firm's success, ect.
Costs $C(a)$

type of agent (v_a, v_y): measures altruism and "greedyness". Only known
to agent, drawn independently

Agent's choice:

$$\max_{a \in A} a(v_a + y v_y) - C(a) + E\left[v_a \mid a, y\right] - E\left[v_y \mid a, y\right]$$

y: extrinsic incentive, can be negative
$E\left[v_a \mid a, y\right]$, $E\left[v_y \mid a, y\right]$: Observers' ex-post expectations of v_a, v_y

Assume that $C(a)$ and distribution of (v_a, v_y) such that optimal choice characterized by FOC:

$$C'(a) = v_a + yv_y + \frac{\partial E[v_a | a, y]}{\partial a} - \frac{\partial E[v_y | a, y]}{\partial a}$$

Three different motivations to contribute: altruism, selfishness, and reputation

Since no heterogeneity between agents, the actual choice of action reveals $v_a + yv_y$ of an agent.

For any $a > 0$, an agent contributes at least a iff

$$v_a + yv_y \geq C'(a) - \frac{\partial E[v_a | a, y]}{\partial a} + \frac{\partial E[v_y | a, y]}{\partial a}$$

In (v_a, v_y) space, the slope of the boundary of this condition is $-\frac{1}{y}$

If $y = 0$ the boundary of this condition is vertical, and the choice of action fully reveals v_a of agent

3 effects of increase of y:

Without reputation: more agents choose more than a - normal incentive effect

"High contributors" are on average less social than before, but also low contributors are: sign of $\frac{\partial E[v_a|a,y]}{\partial y}$ unclear

High contributors are on average more selfish than before, but low contributors are less: sign of $\frac{\partial E[v_y|a,y]}{\partial y}$ positive, less agents choose more than a

Overall effect can be such that with incentives less agents choose weakly more than a - explicit incentives are counterproductive

Experimental test - Ariely et al (2009)

Real effort task for charity

2x2 design: public versus private, monetary incentives versus none

Results:
Without incentives, donataions larger in public than in private treatment
Monetary incentives (small) negative effect in public, but positive effect in private treatment
Highest effort: Public without incentives $\Rightarrow$

Crowding out of incentives

Same result in field study