Intrinsic Motivation and Crowding Out

- Richard Titmuss (1971) found that the US where blood-donors are paid had lower quality blood supplied than UK where it was based on voluntary donation.

- Subsequent experiments (surveyed by Frey and Jegen, 2000 and Fehr and Gachter, 2001) provides evidence for crowding out of intrinsic motivation if monetary incentives are provided.

- "Pay enough or don't pay at all" (experiments by Gneezy and Rustichini (2000))

- Arrow (1972) and Solow (1971) in their surveys of Titmuss thought that doing something for money simply expands the choice set - how can that hurt?
• If you derive intrinsic motivation, and get paid, just add them up

• Also, can donate the money back to your favorite charity

• Can we explain this using the standard economic framework?

• Frey and Oberholzer-Gee (1997) and Gann (2001): giving blood and selling blood are two distinct activity and introducing money transforms the former into the latter

• Frey (1997): preferences change if money is introduced

• Arbitrary, since does not look at consequences.
• Seabright (2002) and Benabou and Tirole (2005): doing something for free signals your type and you build a reputation

• Here we provide an alternative, simpler story, based on heterogeneity in motivation and unobservability of quality.

- A group of agents all care equally about a project
- Benefit from completion normalized to 0
- Only one person can do it
- The "leader" sequentially approaches agents
- If one turns down, there is delay and the project is sent to another agent
- Cost of undertaking the project to an individual $c$
- Private information
• CDF of cost in the population $F(c)$ is common knowledge

• If an agent who is approached turns down, there is a fixed delay cost of $\delta$ each time

• No discounting of future

• Let $\hat{c}$ be the cost threshold such that someone with $c \leq \hat{c}$ will agree (to be determined endogenously)

• Then total expected delay cost is given by the recursive equation

$$D = F(\hat{c}) \times 0 + (1 - F(\hat{c})) \times (\delta + D)$$

• This yields

$$D = \frac{1 - F(\hat{c})}{F(\hat{c})} \delta$$
• A person approached who turns down will expect delay costs to be

\[ \delta + D = \frac{\delta}{F(\hat{\delta})}. \]

• Without payment the relevant cost threshold is

\[ \hat{c}_0 = \frac{\delta}{F(\hat{c}_0)}. \]

• The payoff to each (non-contributing) agent (including leader) is

\[ -D = -\frac{1 - F(\hat{c}_0)}{F(\hat{c}_0)} \delta = \delta - \frac{\delta}{F(\hat{c}_0)} \]

• Suppose there is a monetary incentive of \( w \) to the agent who agrees
• This raises the threshold cost $c$ to

$$\hat{c}_w = w + \frac{\delta}{F(\hat{c}_w)}$$

• Typically $\hat{c}_w > \hat{c}_0$ as $F(.)$ is increasing
Example: Uniform distribution

- Let \( c \in [0, \bar{c}] \) and \( f(c) = \frac{1}{c} \)

- \( F(c) = \int_{0}^{c} f(c) \, dc = \frac{c}{\bar{c}} \)

- Verify: \( \hat{c}_0 = \frac{\delta}{F(\hat{c}_0)} = \frac{\delta}{\frac{\hat{c}_0}{\bar{c}}} \)

- This solves:

\[
\hat{c}_0 = \sqrt{\delta \bar{c}}.
\]

- Also,

\[
\hat{c}_w = w + \frac{\delta \bar{c}}{\hat{c}_w}
\]

- Or,

\[
\hat{c}_w - \frac{\delta \bar{c}}{\hat{c}_w} = w
\]
• Check that LHS (say, \( g(\hat{c}_w) \)) is increasing in \( \hat{c}_w \):
  slope \( 1 + \frac{\delta c}{\hat{c}_w^2} \)

• Also, concave: second derivative

\[
\frac{-2\delta c}{\hat{c}_w^3}
\]

• Therefore, has to be increasing in \( w \) (see Figure 1)
• The payoff of the "leader" is

\[- \frac{1 - F(\hat{c}_w)}{F(\hat{c}_w)} \delta - w.\]

• For \( w = 0 \) we naturally get the same expression as above.

• But

\[ w = \hat{c}_w - \frac{\delta}{F(\hat{c}_w)} \]

• Substituting, the payoff of the "leader" is:

\[ \delta - \hat{c}_w \]

• This is decreasing in \( \hat{c}_w \) and so the optimal \( w \) is 0.
• Suppose the fee is collectively raised and so the leader does not have to put his private weight on the cost (i.e., 1)

• Now the person who agrees gets

\[ \hat{c}_w = w\left(1 - \frac{1}{n}\right) + \frac{\delta}{F(\hat{c}_w)} \]

• This gives

\[ w = \frac{n}{n-1} \left( \hat{c}_w - \frac{\delta}{F(\hat{c}_w)} \right) \]

• Now the leader gets

\[ \frac{-1 - F(\hat{c}_w)}{F(\hat{c}_w)} \delta - \frac{w}{n} = \left[ \delta - \frac{\delta}{F(\hat{c}_w)} \right] + \left[ -\frac{1}{n-1} \left( \hat{c}_w - \frac{\delta}{F(\hat{c}_w)} \right) \right]. \]
• For $w \to 0$, the first term goes to $\delta - \frac{\delta}{F(\hat{c}_0)}$ from above (as it is decreasing in $\hat{c}_w$)

• The second term goes to 0 but is always negative

• So in this case the leader will choose some $w > 0$. 
• Intuition

  – payment of money raises the acceptance rate \( F(\hat{c}_w) > F(\hat{c}_0) \)

  – but that means it’s less costly to turn down a request

  – that means the wage payment has to go up to offset this

• The question, is why does it go up so much so that the leader does not want to do it?

• Payoff of each agent has the term \(-D\)

• The agent who agrees to do it has an added term \( w - \hat{c}_w = - (\delta + D) \)
• For the leader, the payoff has an added term 
\[ w = \delta + D - \hat{c}_w \]

• The leader has to compensate the agent for his private cost \( \hat{c}_w \) out of his pocket, which is increasing in \( w \) and this is not worthwhile.

• Since the agent who agrees internalizes the change in \( D \) and so does the editor, the result of introducing fees on \( D \) cancels out.

• However, if there is cost sharing then the result is not as sharp.

• In general, raising \( w \) does raise the "number" of people who would like to agree.

• This may not hold more generally.
Benabou and Tirole (AER December 2006)  
"Incentives and Pro-Social Behaviour"

- Key observation: rewards change the pool of participants

- You care about your reputation

- However, if money is involved then this is diluted

- Might case some people to drop out

- Participation decision is binary: $a \in \{0, 1\}$

- Costs $C(0) = 0$ and $C(1) = c$

- Motivational reward $v_a \geq 0$
• Monetary reward from income $y : v_y y$ where $v_y \geq 0$

• These parameters $(v_a, v_y)$ are private information

• An individual participates if

$$v_a + yv_y - c + R(y) \geq 0$$

• $R(y)$ captures reputational concerns

$$R(y) \equiv \mu_a \{E(v_a|1, y) - E(v_a|0, y)\} - \mu_y \{E(v_y|1, y) - E(v_y|0, y)\}$$

• The first term captures gains from being known as a "good" citizen and the second losses from being known as "money-minded".

• Assume $v_a$ and $v_y$ are independent so that $f(v_a, v_y) = g(v_a)h(v_y)$
• Assume $\mu_a$ and $\mu_y$ are fixed

• First consider no rewards: $y = 0$

• Then agent participates if and only if

\[ v_a \geq c_a - R(0) \equiv v_a^* \]

• Nothing is learnt about $v_y$ through participation

• A threshold for $v_a$ is learnt

• To determine this let

\[ M(v_a) \equiv E(\tilde{v}_a | \tilde{v}_a \geq v_a) = \frac{\int_{v_a}^{\infty}vg(v)dv}{\int_{v_a}^{\infty}g(v)dv} \]
• Similarly, let

\[ N(v_a) \equiv E(\tilde{v}_a | \tilde{v}_a \leq v_a) = \frac{\int_{-\infty}^{v_a} vg(v) dv}{\int_{-\infty}^{v_a} g(v) dv}. \]

• \( M \) is the average value of \( v_a \) for those who participate

• \( N \) is the average value of \( v_a \) for those who do not participate

• \( M \) is honour from participation and \( N \) is stigma from non-participation

• The cutoff from unpaid participation is then defined as a solution to \( \Phi(v^*_a) = c \) where \( \Phi(v_a) \equiv v_a + \mu_a [M(v_a) - N(v_a)] \), i.e.,

\[ v_a + \mu_a [M(v_a) - N(v_a)] = c. \]
• For the uniform distribution

\[ M(v_a) - N(v_a) = \frac{1}{2} \text{ for } g(v_a) = 1 \text{ on } [0, 1] \]

• Proof:

\[
\frac{\int_{v_a}^1 v g(v) \, dv}{\int_{v_a}^1 g(v) \, dv} - \frac{\int_0^{v_a} v g(v) \, dv}{\int_0^{v_a} g(v) \, dv} = \frac{\int_{v_a}^1 v \, dv}{\int_{v_a}^1 g(v) \, dv} - \frac{\int_0^{v_a} v \, dv}{\int_0^{v_a} g(v) \, dv} \\
= \frac{\left[ \frac{v^2}{2} \right]_1^{v_a}}{[v]_{v_a}^1} - \frac{\left[ \frac{v^2}{2} \right]_0^{v_a}}{[v]_0^{v_a}} \\
= \frac{1}{2} \left( \frac{1 - v_a^2}{1 - v_a} - \frac{v_a^2}{1 - v_a} \right) \\
= \frac{1}{2} \left( 1 + v_a - v_a \right)
\]

• More generally, for \( g(v_a) = (\alpha + 1) v_a^\alpha \) on \([0, 1]\) with \( \alpha > -1 \)

\[
M(v_a) - N(v_a) = \frac{1 + \alpha}{2 + \alpha} \frac{1 - v_a}{1 - v_a^{1+\alpha}}
\]
• This is increasing in $v_a$ when $\alpha > 0$ and decreasing when $-1 < \alpha < 0$.

• Key Result: Assume that $\Phi'(.) \geq 0$. Then if $\mu_y = 0$ or if $v_a$ and $v_y$ are independent, the introduction of reward lowers net reputational value of participation: $R(y) < R(0)$ for all $y > 0$.

• The assumption $\Phi'(.) \geq 0$ ensures uniqueness of equilibrium.

• If you introduce reward, $v_a + yv_y - c + R(y) \geq 0$

• Intuition: Two reputational effects
– Reward attracts greedy types: $E(v_y|1, y) > E(v_y|0, y)$

– Reward repels good types

• Overall participation will increase or decrease depending on the weights of these two effects

• Note here the quality of participation is the same for all types so therefore cannot explain Titmuss’ findings
Extrinsic Vs. Intrinsic Incentives (Benabou-Tirole, 2003): To be added.
Figure 1