Supplementary Material
Costly incentives design from an institutional perspective: cooperation, sustainability, and affluence

Xin Zhou*, Adam Belloum, Michael H. Lees, Tom van Engers, Cees de Laat
University of Amsterdam, Amsterdam, 1098XH, Netherlands

Supplementary part 1. Population equilibrium
To predict the effect of incentive mechanism, evolutionary game theory which describes the population of agents engaging in pairwise interaction, has been generally accepted as a common framework to model and interpret the evolution of cooperation in a social dilemma. This part shows the derivation process of the analytical results of the population in equilibrium.

Assume a finite population of size $N$, $M$ of them participate a market play PDG, two types of strategies for cooperation $C$ and defection $D$ are well mixed. Let $\pi(C)$ denote the expected payoff of strategy $C$, $\pi(D)$ donate that of strategy $D$, and $\bar{\pi}(x)$ denote the average payoff of the whole population:

$$\pi(C) = (1 - c_0 + R_{CC})x + (-T - c_0 + R_{CD})y$$

$$\pi(D) = (T - c_0 - F_{CD})x + (0 - c_0 - F_{DD})y$$

$$\bar{\pi}(x) = \pi(C)x + \pi(D)y$$

The replicator dynamics of cooperators and defectors is:

$$\dot{x} = x(1 - x)[\pi(C) - \pi(D)].$$

With the equations (1), (2), (3), and (4), let $\dot{x} = 0$,

$$(x^2 - x)[x + xR_{CC} + (1 - x)R_{CD} + xF_{CD} + (1 - x)F_{DD} - T] = 0$$

Thus, the fixed points are $x^* = 0$, $x^* = 1$, and $x^* = \frac{-F_{DD} + T - R_{CD}}{1 + R_{CC} - R_{CD} + F_{CD} - F_{DD}}$.

We next discuss if the fixed point is the Nash equilibrium (NE) and satisfy the requirement of being as the evolutionary stable strategy (ESS).

Let $s^* = (x^*, 1 - x^*)$, $s = (p, 1 - p)$, $s \neq s^*$, an ESS can be defined as a mixed strategy $s^*$, such that for any strategy $s$ and any sufficient small $\varepsilon > 0$,

$$\pi[s^*, (1 - \varepsilon)s^* + \varepsilon s] > \pi[s, (1 - \varepsilon)s^* + \varepsilon s].$$

Using the linearity in probability of expected payoffs reduces 6 to:

$$(1 - \varepsilon)\pi(s^*, s^*) + \varepsilon\pi(s^*, s) > (1 - \varepsilon)\pi(s, s^*) + \varepsilon\pi(s, s).$$

If for all small $\varepsilon > 0$ and for all $s$,

$$\pi(s^*, s) \geq \pi(s, s^*),$$

then $s^*$ is a symmetric NE. Further, if

$$\pi(s^*, s) > \pi(s, s)$$

whenever $\pi(s^*, s^*) = \pi(s, s^*)$, $s^*$ is an ESS [1].
1) fixed point at $x^* = 0, s^* = (0,1), s = (p,1-p) (0 < p \leq 1)$, we have:

$$\pi(s^*, s^*) = -c_0 - F_{DD}$$  \hspace{1cm} (8a)
$$\pi(s, s^*) = p(-T - c_0 + R_{CD}) + (1 - p)(0 - c_0 - F_{DD})$$  \hspace{1cm} (8b)
$$\pi(s^*, s) = p(T - c_0 - F_{CD}) + (1 - p)(0 - c_0 - F_{DD})$$  \hspace{1cm} (9a)
$$\pi(s, s) = p^2(1 - c_0 + R_{CC}) + p(1 - p)(-T - c_0 + R_{CD})$$
$$+ p(1 - p)(T - c_0 - F_{CD}) + (1 - p)^2(0 - c_0 - F_{DD})$$  \hspace{1cm} (9b)

By $8a - 8b$, we have

$$8a - 8b = -p(-T - c_0 + R_{CD}) + p(0 - c_0 - F_{DD})$$
$$= p(T - F_{DD} - R_{CD}),$$  \hspace{1cm} (8c)

and by $9a - 9b$, we have

$$9a - 9b = p^2(T - F_{CD} - R_{CC}) + p(1 - p)(-F_{DD} + T - R_{CD}).$$  \hspace{1cm} (9c)

The Table 1 shows the requirements for $x^* = 0$ being a NE or an ESS under different incentive policies.

<table>
<thead>
<tr>
<th>Incentive Policy</th>
<th>Scale of parameters</th>
<th>NE</th>
<th>ESS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reward</strong></td>
<td>$0 &lt; R_{CC} + R_{CD}, F_{CD} = 0$</td>
<td>When $T \geq R_{CD}, 8c \geq 0$</td>
<td>When $T &gt; R_{CD}, 9c &gt; 0$</td>
</tr>
<tr>
<td><strong>Punishment</strong></td>
<td>$R_{CD} = R_{DD} = 0, 0 &lt; F_{CD} + F_{DD}$</td>
<td>When $T \geq F_{DD}, 8c \geq 0$</td>
<td>When $T \geq F_{CD}, 9c &gt; 0$</td>
</tr>
<tr>
<td><strong>Mixed</strong></td>
<td>$R_{CD} + R_{CC} \neq 0, F_{CD} + F_{DD} \neq 0$</td>
<td>When $T \geq F_{CD} + F_{DD}, 8c \geq 0$</td>
<td>When $T \geq \max[R_{CD} + F_{DD}, R_{CC} + F_{CD}], 9c &gt; 0$</td>
</tr>
</tbody>
</table>

2) Consider the fixed point at $x^* = 1, s^* = (1,0), s = (1 - p, p) (0 < p \leq 1)$, we have:

$$\pi(s^*, s^*) = 1 - c_0 + R_{CC}$$  \hspace{1cm} (8d)
$$\pi(s, s^*) = (1 - p)(1 - c_0 + R_{CC}) + p(T - c_0 - F_{CD})$$  \hspace{1cm} (8e)
$$\pi(s^*, s) = (1 - p)(1 - c_0 + R_{CC}) + p(-T - c_0 + R_{CD})$$  \hspace{1cm} (9d)
$$\pi(s, s) = (1 - p)^2(1 - c_0 + R_{CC}) + p(1 - p)(-T - c_0 + R_{CD})$$
$$+ p(1 - p)(T - c_0 - F_{CD}) + p^2(0 - c_0 - F_{DD})$$  \hspace{1cm} (9e)

By $8d - 8e$, we have

$$8d - 8e = p(1 - c_0 + R_{CC}) - p(T - c_0 - F_{CD})$$
$$= p(1 - T + R_{CC} + F_{CD}),$$  \hspace{1cm} (8f)

and by $9d - 9e$, we have

$$9d - 9e = p(1 - p)(1 - c_0 + R_{CC}) - p(1 - p)(T - c_0 - F_{CD})$$
$$+ p^2(-T - c_0 + R_{CD}) - p^2(0 - c_0 - F_{DD})$$  \hspace{1cm} (9f)
$$= p(1 - p)(1 - T + R_{CC} + F_{CD}) + p^2(-T + R_{CD} + F_{DD}).$$

The Table 2 shows the requirements for $x^* = 1$ being a NE or an ESS under different incentive policies.
3) Consider the fixed point at \( x^* = \frac{T - R_{CD} - F_{DD}}{1 - R_{CC} - R_{CD} - F_{CD} - F_{DD}} = \frac{B}{A} \). Let \( q = \frac{B}{A}, s^* = (q, 1 - q), s = (p, 1 - p) \) \((0 \leq p \leq 1, p \neq q)\), we have:

\[
\pi(s', s^*) = q^2(1 - c_0 + R_{CC}) + q(1 - q)(-T - c_0 - R_{CD}) + q(1 - q)(T - c_0 - F_{CD}) + (1 - q)^2(0 - c_0 - F_{DD})
\]

\[
= q(1 - q)(T - c_0 - F_{CD}) + (1 - q)^2(0 - c_0 - F_{DD})
\]

(8g)

\[
\pi(s, s^*) = pq(1 - c_0 + R_{CC}) + p(1 - q)(-T - c_0 + R_{CD}) + p(1 - q)(T - c_0 - F_{CD}) + (1 - q)(0 - c_0 - F_{DD})
\]

\[
= p(1 - q)(T - c_0 - F_{CD}) + (1 - q)(0 - c_0 - F_{DD})
\]

(8h)

\[
\pi(s, s) = p^2(1 - c_0 + R_{CC}) + p(1 - p)(-T - c_0 + R_{CD}) + p(1 - p)(T - c_0 - F_{CD}) + (1 - p)^2(0 - c_0 - F_{DD})
\]

(9g)

By 8g - 8h, we have

\[
8g - 8h = q(q - p)(1 - c_0 + R_{CC} - T + c_0 + F_{CD}) + (1 - q)(q - p)(-T - c_0 + R_{CD} + c_0 + F_{DD})
\]

\[
= q(q - p)(1 - T + R_{CC} + F_{CD}) + (1 - q)(q - p)(-T + R_{CD} + F_{DD})
\]

\[
= (q - p)[q(1 + R_{CC} + F_{CD} + R_{CD} - F_{DD}) - T + R_{CD} + F_{DD}]
\]

\[
= (q - p)[q(1 + R_{CC} + F_{CD} + R_{CD} - F_{DD}) - T + R_{CD} + F_{DD}]
\]

(8i)

and by 9g - 9h, we have

\[
9g - 9h = p(q - p)(1 - c_0 + R_{CC} - T + c_0 + F_{CD}) + (1 - p)(q - p)(-T - c_0 + R_{CD} + c_0 + F_{DD})
\]

\[
= p(q - p)(1 - T + R_{CC} + F_{CD}) + (1 - p)(q - p)(-T + R_{CD} + F_{DD})
\]

\[
= (q - p)[p(1 + R_{CC} + F_{CD} - R_{CD} - F_{DD}) - T + R_{CD} + F_{DD}]
\]

\[
= (q - p)[p(1 + R_{CC} + F_{CD} - R_{CD} - F_{DD}) - T + R_{CD} + F_{DD}]
\]

(9i)

The Table 3 shows the requirements for \( x^* = q \) being a NE or an ESS under different incentive policies.

### Supplementary table 2. NE and ESS analysis under different conditions when \( x^* = 1 \)

<table>
<thead>
<tr>
<th>Incentive Policy</th>
<th>Scale of parameters</th>
<th>NE</th>
<th>ESS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reward</strong></td>
<td>( 0 &lt; R_{CC} + R_{CD}, F_{CD} = ) ( F_{DD} = 0 )</td>
<td>When ( R_{CC} \geq T - 1, 8f \geq 0 )</td>
<td>When ( R_{CC} \geq T, 9f &gt; 0 )</td>
</tr>
<tr>
<td><strong>Punishment</strong></td>
<td>( R_{CD} = R_{DD} = 0, 0 &lt; ) ( F_{CD} + F_{DD} )</td>
<td>When ( F_{CD} \geq T - 1, 8f \geq 0 )</td>
<td>When ( F_{DD} &gt; T, 9f &gt; 0 )</td>
</tr>
<tr>
<td><strong>Mixed incentives</strong></td>
<td>( R_{CD} + R_{CC} \neq 0, F_{CD} + ) ( F_{DD} \neq 0 )</td>
<td>When ( R_{CC} + F_{CD} \geq T - 1, 8f \geq 0 )</td>
<td>When ( \min[R_{CD} + F_{DD} + R_{CC} + F_{CD}] \geq T, 9f &gt; 0 )</td>
</tr>
</tbody>
</table>

### Supplementary table 3. NE and ESS analysis under different conditions when \( x^* = q \)

<table>
<thead>
<tr>
<th>Incentive Policy</th>
<th>Scale of parameters</th>
<th>NE</th>
<th>ESS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reward</strong></td>
<td>( 0 &lt; R_{CC} + R_{CD}, F_{CD} = ) ( F_{DD} = 0 )</td>
<td>( 8i = 0 ), thus always hold</td>
<td>( 9i &lt; 0 ), thus the ESS is not hold</td>
</tr>
<tr>
<td><strong>Punishment</strong></td>
<td>( R_{CD} = R_{DD} = 0, 0 &lt; ) ( F_{CD} + F_{DD} )</td>
<td>( 8i = 0 ), thus always hold</td>
<td>( 9i &lt; 0 ), thus the ESS is not hold</td>
</tr>
<tr>
<td><strong>Mixed incentives</strong></td>
<td>( R_{CD} + R_{CC} \neq 0, F_{CD} + ) ( F_{DD} \neq 0 )</td>
<td>( 8i = 0 ), thus always hold</td>
<td>( 9i &lt; 0 ), thus the ESS is not hold</td>
</tr>
</tbody>
</table>
Supplementary part 2. Rate of $R_{CC}$ ($F_{DD}$) in $R_{CC} + F_{CD}$ ($R_{CD} + F_{DD}$)

Let $\alpha$ be the rate, $k = R_{CC} + F_{CD}$, to minimize the difference between reward($R_{CD} + R_{CC}$) and punishment ($F_{CD} + F_{DD}$), meanwhile satisfy the constraints are $R_{CD} \geq R_{CC}, F_{CD} \geq F_{DD}$, we have:

$$\min_{\alpha} \quad k\alpha + (k + 1)\alpha - [k(1 - \alpha) + (1 + k)\alpha]$$

s.t. $(1 - \alpha)2 \geq 3\alpha$

$(1 - \alpha)1 \geq 4\alpha$

(10)

The solution is $\alpha = 0.2$. Consequently, these four parameters are set as shown in Table 4.

Supplementary table 4. Mixed incentives setup

<table>
<thead>
<tr>
<th>$R_{CC} + F_{CD}$</th>
<th>$R_{CC}$</th>
<th>$F_{CD}$</th>
<th>$R_{CD} + F_{DD}$</th>
<th>$R_{CD}$</th>
<th>$F_{DD}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.2</td>
<td>0.8</td>
<td>2</td>
<td>1.6</td>
<td>0.4</td>
</tr>
<tr>
<td>1.25</td>
<td>0.25</td>
<td>1</td>
<td>2.25</td>
<td>1.8</td>
<td>0.45</td>
</tr>
<tr>
<td>1.5</td>
<td>0.3</td>
<td>1.2</td>
<td>2.5</td>
<td>2</td>
<td>0.5</td>
</tr>
<tr>
<td>1.75</td>
<td>0.35</td>
<td>1.4</td>
<td>2.75</td>
<td>2.2</td>
<td>0.55</td>
</tr>
<tr>
<td>2</td>
<td>0.4</td>
<td>1.6</td>
<td>3</td>
<td>2.4</td>
<td>0.6</td>
</tr>
<tr>
<td>2.25</td>
<td>0.45</td>
<td>1.8</td>
<td>3.25</td>
<td>2.6</td>
<td>0.65</td>
</tr>
<tr>
<td>2.5</td>
<td>0.5</td>
<td>2</td>
<td>3.5</td>
<td>2.8</td>
<td>0.7</td>
</tr>
<tr>
<td>2.75</td>
<td>0.55</td>
<td>2.2</td>
<td>3.75</td>
<td>3</td>
<td>0.75</td>
</tr>
<tr>
<td>3</td>
<td>0.6</td>
<td>2.4</td>
<td>4</td>
<td>3.2</td>
<td>0.8</td>
</tr>
</tbody>
</table>

Supplementary part 3. Accumulated wealth of the third-party

From equation 4, we have

$$\dot{x} = \frac{dx}{dt} = (1 + R_{CC} - R_{CD} + F_{CD} - F_{DD})x^3$$

$$+ (2R_{CD} + 2F_{DD} - F_{CD} - R_{CC} - T - 1)x^2$$

$$- (R_{CD} + F_{DD} - T)x$$

(11)

Let $a = 1 + R_{CC} - R_{CD} + F_{CD} - F_{DD}, b = 2R_{CD} + 2F_{DD} - F_{CD} - R_{CC} - T - 1, c = T - R_{CD} - F_{DD}$, we have:

$$\frac{dx}{dt} = ax^3 + bx^2 + cx$$

(12)

We can then solve the differential equation:

$$\int \frac{1}{ax^3 + bx^2 + cx} \, dx = \int 1 \, dt$$

$$= \int \frac{1}{x(ax^2 + bx + c)} \, dx$$

$$= \int \frac{1}{cx} + \frac{-ax - b}{c(ax^2 + bx + c)} \, dx$$

$$= \frac{1}{c} \int \frac{-dx}{x} + \frac{a}{c} \int \frac{x}{ax^2 + bx + c} \, dx - \frac{b}{c} \int \frac{1}{ax^2 + bx + c} \, dx$$

$$= \frac{1}{c} \ln x - \frac{a}{c} \left( \frac{\ln(ax^2 + bx + c)}{2a} - \frac{b}{a\sqrt{4ac - b^2}} \tan^{-1} \left( \frac{2ax + b}{\sqrt{4ac - b^2}} \right) \right)$$

$$- \frac{b}{c} \frac{2}{a\sqrt{4ac - b^2}} \tan^{-1} \left( \frac{2ax + b}{\sqrt{4ac - b^2}} \right) + C$$

$$= t$$

$x^{(t)}$ is the inverse function of 13, let $x^{(t)} := F(t, a, b, c)$. The wealth of the third party at time step $t$ is shown as 2.3.
\[ W_T^{(i)} = M^{(i)} \left( c_0 + x^{(i)}(1-x^{(i)})F_{CD} + (1-x^{(i)})^2F_{DD} - (x^{(i)})^2R_{CC} - x^{(i)}(1-x^{(i)})R_{CD} \right. \]
\[
- \alpha \left( x^{(i)}(1-x^{(i)})F_{CD} + (1-x^{(i)})^2F_{DD} \right) \]
\[
= M^{(i)} \left( (x^{(i)})^2 \left( (1-\alpha)(F_{DD} - F_{CD}) - R_{CC} + R_{CD} \right) + x^{(i)} \left( (1-\alpha)(F_{CD} - 2F_{DD}) - R_{CD} \right) \right. \]
\[
+ c_0 + (1-\alpha)F_{DD} \right) \]
\[
= M^{(i)} \left( F^2(t,a,b,c) \left( (1-\alpha)(F_{DD} - F_{CD}) - R_{CC} + R_{CD} \right) \right. \]
\[
+ F(t,a,b,c) \left( (1-\alpha)(F_{CD} - 2F_{DD}) - R_{CD} \right) + c_0 + (1-\alpha)F_{DD} \right) \]

Thus,

\[ W_T = \int W_T^{(i)} \, dt \]
\[
= \left( (1-\alpha)(F_{DD} - F_{CD}) - R_{CC} + R_{CD} \right) \int M^{(i)} F^2(t,a,b,c) \, dt \]
\[
+ \left( (1-\alpha)(F_{CD} - 2F_{DD}) - R_{CD} \right) \int M^{(i)} F(t,a,b,c) \, dt \]
\[
+ (c_0 + (1-\alpha)F_{DD}) \int M^{(i)} \, dt \]

**References**