Supplementary Information

1 Bribery Game

1.1 Game Description

- Public Goods Game variant with centralized institutional punishment; the central authority is randomly selected.
- All players, including the central authority, remain anonymous to each other for the whole duration of the game.
- All players pay the same, fixed amount of tax out of their total endowment at the beginning of each period. The collected tax is transferred to the Punishment Endowment which is used by the central authority to punish players in each period.
- Players can choose between three actions (or a combination of those actions) with regard to the endowment that is equally distributed among all players at the beginning of each period:
  - Contribute to the public good
  - Make payment to the central authority (i.e., bribe)
  - Keep money for self
- The central authority should choose only one of three possible actions for every player after observing players' decisions at the beginning of each period:
  - Accept bribe (if any bribe is offered)
  - Punish
  - Do nothing

1.2 Experimental Conditions

Five different conditions were used to run the Bribery Game experiments. In each experimental session, participants played four randomly chosen conditions. These conditions are:

- Control: Public goods game without bribery, with centralized institutional punishment.
- Bribery Game: Public goods game that allows players to bribe the central authority.
- Bribery Game with Partial Transparency: Participants can observe the central authority's contribution to the public good at the end of each period of the Bribery Game.

- Bribery Game with Full Transparency: Participants can observe all players' actions (i.e., every player's contribution to the public good and offered bribe to the central authority, and also the central authority's contribution to the public good and decisions for every player) at the end of each period of the Bribery Game.

- Bribery Game with Maximum Central Authority Contribution: The central authority does not have the option of choosing his own contribution to the public good at the beginning of each period of the Bribery Game, and his total endowment (after paying the mandatory tax) is automatically transferred to the public good.

1.3 Theoretical Framework

Variables:

e: total endowment ($e = 12$ in BG experiments)
t: tax ($t = 2$ in BG experiments)
m: the public good multiplier ($m \in \{1.2, 1.5, 1.8, 2.1, 2.4, 3, 3.6, 4.2\}$ in BG experiments)
n: number of players ($n \in \{4, 5, 6, 7\}$ in BG experiments)
r: the punishment multiplier ($r \in \{1, 3\}$ in BG experiments)
g: contribution to the public good
b: bribe
A: the central authority
d_{Ai}: punishment assigned by the central authority to player i

1.3.1 Condition 0 (Control)

Condition 0: Institutional punishment public goods game. Central authority can choose his own contribution to the public pool. No bribery.

Monetary payoffs:

$$\pi_i = e - t - g_i + \frac{m}{n}\sum_{j=1}^{n} g_j - rd_{Ai} \text{ if } i \neq A,\ i \in \{1, ..., n\}$$

$$\text{Central Authority: } \pi_A = e - t - g_A + \frac{m}{n}\sum_{j=1}^{n} g_j,\ A \in \{1, ..., n\}$$

<u>*Standard Model*</u> (with the assumption that individuals simply maximize their monetary payoffs)

$$\frac{\partial \pi_i}{\partial g_i} = \frac{m}{n} - 1 < 0 \ (\frac{m}{n} \in \{0.3, 0.6\} \text{ in BG experiments})$$

$g_i = 0$ is the dominant strategy for all $i \in \{1, ..., n\}$.
Equilibria: $g_i = 0$ and any d_A (vector of punishments) for all $i \in \{1, ..., n\}$ and $0 \leq \sum_{i \neq A} d_{Ai} \leq nt$.

<u>*Inequity Aversion Model*</u>

Utility function:

$$u_i(\pi) = \pi_i - \alpha_i \frac{1}{n-1} \sum_{j \neq i} max\{\pi_j - \pi_i, 0\} - \beta_i \frac{1}{n-1} \sum_{j \neq i} max\{\pi_i - \pi_j, 0\} \text{ where } \pi = (\pi_1, ..., \pi_n)$$

α: the disadvantageous inequality multiplier
β: the advantageous inequality multiplier

assumptions: $\beta_i \leq \alpha_i$ and $0 \leq \beta_i < 1$ (Fehr and Schmidt 1999)

Proposition:

There are many equilibria where $g_i = g \in [0, min\{e-t, rnt\}]$ for all $i \in \{1, ..., n\}$ if $\frac{m}{n} + \beta_A \geq 1$; if individual K ($K \in \{1, ..., n\}$ and $K \neq A$) deviates to $g_K < g$, he would receive a punishment of $d_{AK} = \frac{g - g_K}{r}$.

Proof:

If $g_i = g$:

$$\pi^* = \pi_i = e - t + (m-1)g$$
$$u^* = u_i = \pi^*$$

Lets assume individual K ($K \neq A$) deviates to $g_K < g$:

$$\pi_K = e - t - g_K + \frac{m}{n}[(n-1)g + g_K] - rd_{AK}$$
$$\pi_i = e - t - g + \frac{m}{n}[(n-1)g + g_K] \text{ if } i \neq K$$

$$\pi_K = \pi_i \Rightarrow d^*_{AK} = \frac{g - g_K}{r}$$

When $d_{AK} = d^*_{AK}$, $\pi^* > \pi_K$ and individual K would have no incentive to deviate to g_K.

If the central authority deviates to $g_A < g$:

$$u^* - u_A = -(g - g_A)[1 - \frac{m}{n} - \beta_A]$$

Therefore, the central authority would have no incentive to deviate if $\frac{m}{n} + \beta_A \geq 1$.

In addition, the central authority has no incentive to deviate from d^*_{AK} when individual K contributes $g_K < g$:

If $d_{AK} = d^*_{AK} + \epsilon$, then $u^* = u_A + \frac{\beta_A}{n-1} r\epsilon$.
If $d_{AK} = d^*_{AK} - \epsilon$, then $u^* = u_A + \frac{\alpha_A}{n-1} r\epsilon$.

Also,

$$\frac{g - g_K}{r} \leq nt$$

$$g_K = 0 \Rightarrow g \leq rnt$$

The condition $g \leq rnt$ guarantees that the central authority has enough resources to punish any deviation g_K.

1.3.2 Conditions 1, 2, and 3

Condition 1: Bribery Game

Condition 2: Bribery Game with partial transparency. Central authority can choose his own contribution to the public pool. The value of this contribution is visible to all players.

Condition 3: Bribery Game with full transparency. Central authority can choose his own contribution to the public pool. The value of this contribution, all player contributions, and all player bribes are visible to all players (anonymized).

$$\pi_i = e - t - g_i - b_i + \frac{m}{n}\sum_{j=1}^{n} g_j - rd_{Ai} \text{ if } i \neq A$$

$$\pi_A = e - t - g_A + \frac{m}{n}\sum_{j=1}^{n} g_j + \sum_{j \neq A} b_j$$

Proposition:

There are many equilibria where $g_i = g \in [0, min\{e-t, rnt\}]$ for all $i \in \{1, ..., n\}$ and $b_i = 0$ for all $i \neq A$ if $\frac{m}{n} + \beta_A \geq 1$;
if individual K deviates to $g_K < g$, the central authority would assign a punishment of $d_{AK} = \dfrac{g - g_K}{r}$.

Proof:

If $g_i = g$ for all $i \in \{1, ..., n\}$ and $b_i = 0$ for all $i \neq A$:

$$\pi^* = \pi_i = e - t + (m-1)g$$

Lets assume individual K ($K \in \{1, ..., n\}$ and $K \neq A$) deviates to $g_K < g$:

$$\pi_K = e - t - g_K - b_K + \frac{m}{n}[(n-1)g + g_K] - rd_{AK}$$

$$\pi_A = e - t - g + \frac{m}{n}[(n-1)g + g_K] + b_K$$

$$\pi_A = \pi_K \Rightarrow 2b_K + rd_{AK} = g - g_K$$

Case 1. $b_K = 0$
The central authority assigns a punishment of $d_{AK} = \dfrac{g - g_K}{r}$ and therefore, individual K would have no incentive to deviate.

Case 2. $d_{AK} = 0$

The central authority does not punish if he is offered a payment of $b_K = \dfrac{g - g_K}{2}$:

$$b_K = \frac{g - g_K}{2} \Rightarrow \pi_K = e - t - \frac{g + g_K}{2} + \frac{m}{n}[(n-1)g + g_K] < \pi^*$$

Therefore, individual K would have no incentive to deviate.

Also,
$d_{AK} \leq nt \Rightarrow g \leq rnt$ if $g_K = 0$

If the central authority deviates to $g_A < g$:

$$u_A = e - t - g_A + \frac{m}{n}[(n-1)g + g_A] - \beta_A(g - g_A)$$

$$u^* > u_A \Rightarrow \frac{m}{n} + \beta_A \geq 1$$

Therefore, the central authority would have no incentive to deviate to $g_A < g$ if $\frac{m}{n} + \beta_A \geq 1$.

1.3.3 Condition 4

Condition 4: Bribery Game with forced leader investment. Central authority is forced to contribute entire endowment to the public pool.

$$\pi_i = e - t - g_i - b_i + \frac{m}{n}\sum_{j=1}^{n} g_j - r d_{Ai}$$

$$\pi_A = \sum_{j \neq A} b_j + \frac{m}{n}\sum_{j=1}^{n} g_j$$

Inequity Aversion Model

Proposition:

There are many equilibria where $g_i = g \in [0, min\{e - t, \frac{rn^2t - 2(e-t)}{n-2}\}]$, $b_i = b \in [\frac{e-t-rnt}{n-2}, \frac{e-t}{n}]$, and $b = \dfrac{e - t - g}{n}$ for all $i \neq A$ given that $0 \leq g_i + b_i \leq e - t$;

if individual K deviates, he would receive a punishment of $d_{AK} = \dfrac{1}{r}[2b + (g - g_K)]$.

Proof:

$$\pi_i = e - t - g - b + \frac{m}{n}[(n-1)g + e - t] \text{ if } i \neq A$$

$$\pi_A = (n-1)b + \frac{m}{n}[(n-1)g + e - t]$$

$$\pi_i = \pi_A \Rightarrow b = \frac{e - t - g}{n}$$

$$g = 0 : b \leq \frac{e-t}{n}$$

$$d_{AK} \leq nt \Rightarrow g \leq \tfrac{rn^2t-2(e-t)}{n-2} \text{ and } b \geq \tfrac{e-t-rnt}{n-2}$$

For individual K:

$$\pi_K = e - t - g_K - b_K + \frac{m}{n}[(n-2)g + g_K + e - t]$$

$$\pi_K > \pi^* \Rightarrow (g+b) - (g_K + b_K) > \frac{m}{n}(g - g_K)$$

Case 1. $g_K = g$ and $b_K < b$:

$$\pi_K = e - t - g - b_K + \frac{m}{n}[(n-1)g + e - t] - rd_{AK}$$

$$\pi_A = (n-2)b + b_K + \frac{m}{n}[(n-1)g + e - t]$$

$$\pi_i = e - t - g - b + \frac{m}{n}[(n-1)g + e - t] = (n-1)b + \frac{m}{n}[(n-1)g + e - t]$$

If Punish:
($b_K = 0$)

$$\pi_K = nb + \frac{m}{n}[(n-1)g + e - t] - rd_{AK}$$

$$\pi_A = (n-2)b + \frac{m}{n}[(n-1)g + e - t]$$

$$\pi_A = \pi_K \Rightarrow nb - rd_{AK} = (n-2)b \Rightarrow rd_{AK} = 2b \Rightarrow d_{AK} = \frac{2b}{r}$$

$$u_{A,P} = (n-2)b + \frac{m}{n}[(n-1)g + e - t] - \alpha_A \frac{n-2}{n-1} b$$

If Accept Bribe:
($d_{AK} = 0$)

$$\pi_K = nb - b_K + \frac{m}{n}[(n-1)g + e - t]$$

$$\pi_A = (n-2)b + b_K + \frac{m}{n}[(n-1)g + e - t]$$

$$u_{A,B} = (n-2)b + b_K + \frac{m}{n}[(n-1)g + e - t] - \alpha_A \frac{n}{n-1}(b - b_K)$$

Punish if:

$$u_{A,P} > u_{A,B} \Rightarrow -\alpha_A \frac{n-2}{n-1} b > b_K - \alpha_A \frac{n}{n-1}(b - b_K) \Rightarrow \frac{b}{b_K} > \frac{n\alpha_A + n - 1}{2\alpha_A}$$

Case 2. $g_K < g$ and $b_K = b$:

$$\pi_K = e - t - g_K - b + \frac{m}{n}[(n-2)g + g_K + e - t] - rd_{AK}$$

$$\pi_A = (n-1)b + \frac{m}{n}[(n-2)g + g_K + e - t]$$

$$\pi_i = e - t - g - b + \frac{m}{n}[(n-2)g + g_K + e - t] = (n-1)b + \frac{m}{n}[(n-2)g + g_K + e - t]$$

If Punish:
($b_K = 0$)

$$\pi_K = e - t - g_K + \frac{m}{n}[(n-2)g + g_K + e - t] - rd_{AK}$$

$$\pi_A = (n-2)b + \frac{m}{n}[(n-2)g + g_K + e - t]$$

$$\pi_A = \pi_K \Rightarrow e - t - g_K - rd_{AK} = (n-2)b \Rightarrow rd_{AK} = e - t - g + (g - g_K) - (n-2)b \Rightarrow d_{AK} = \frac{2b + (g - g_K)}{r}$$

$$u_{A,P} = (n-2)b + \frac{m}{n}[(n-2)g + g_K + e - t] - \alpha_A \frac{n-2}{n-1} b$$

If Accept Bribe:
($d_{AK} = 0$)

$$\pi_K = e - t - g_K - b + \frac{m}{n}[(n-2)g + g_K + e - t] = (n-1)b + g - g_K + \frac{m}{n}[(n-2)g + g_K + e - t]$$

$$\pi_A = (n-1)b + \frac{m}{n}[(n-2)g + g_K + e - t]$$

$$u_{A,B} = (n-1)b + \frac{m}{n}[(n-2)g + g_K + e - t] - \alpha_A \frac{g - g_K}{n-1}$$

Punish if:

$$u_{A,P} > u_{A,B} \Rightarrow (n-2)b - \alpha_A \frac{n-2}{n-1} b > (n-1)b - \alpha_A \frac{g - g_K}{n-1} \Rightarrow g - g_K > (n - 2 + \frac{n-1}{\alpha_A})b$$

Case 3. $g_K < g$ and $b_K > b$:

$$\pi_K = e - t - g_K - b_K + \frac{m}{n}[(n-2)g + g_K + e - t] - rd_{AK}$$

$$\pi_A = (n-2)b + b_K + \frac{m}{n}[(n-2)g + g_K + e - t]$$

$$\pi_i = e - t - g - b + \frac{m}{n}[(n-2)g + g_K + e - t] = (n-1)b + \frac{m}{n}[(n-2)g + g_K + e - t]$$

If Punish:
($b_K = 0$)

$$\pi_K = e - t - g_K + \frac{m}{n}[(n-2)g + g_K + e - t] - rd_{AK}$$

$$\pi_A = (n-2)b + \frac{m}{n}[(n-2)g + g_K + e - t]$$

$$\pi_A = \pi_K \Rightarrow e - t - g_K - rd_{AK} = (n-2)b \Rightarrow d_{AK} = \frac{2b + (g - g_K)}{r}$$

$$u_{A,P} = (n-2)b + \frac{m}{n}[(n-2)g + g_K + e - t] - \alpha_A \frac{n-2}{n-1} b$$

If Accept Bribe:
($d_{AK} = 0$)

$$\pi_K = e - t - g_K - b_K + \frac{m}{n}[(n-2)g + g_K + e - t]$$

$$\pi_A = (n-2)b + b_K + \frac{m}{n}[(n-2)g + g_K + e - t]$$

$u_{A,B} = (n-2)b + b_K + \frac{m}{n}[(n-2)g + g_K + e - t] - \beta_A \frac{n(b_K - b) - (g - g_K)}{n-1}$ if $2(b_K - b) > g - g_K$
$u_{A,B} = (n-2)b + b_K + \frac{m}{n}[(n-2)g + g_K + e - t] - \alpha_A \frac{(g-g_K) - 2(b_K - b)}{n-1} - \beta_A \frac{n-2}{n-1}(b_K - b)$ if $2(b_K - b) < g - g_K$

Punish if:
$b_K(\beta_A n - n + 1) - b[\beta_A n + \alpha_A(n-2)] > \beta_A(g - g_K)$ when $2(b_K - b) > g - g_K$
$\alpha_A(g - g_K) > b_K[n - 1 + 2\alpha_A - \beta_A(n-2)] - b[n\alpha_A - (n-2)\beta_A]$ when $2(b_K - b) < g - g_K$

Case 4. $g_K < g$ and $b_K < b$:

$$\pi_K = e - t - g_K - b_K + \frac{m}{n}[(n-2)g + g_K + e - t] - rd_{AK}$$

$$\pi_A = (n-2)b + b_K + \frac{m}{n}[(n-2)g + g_K + e - t]$$

$$\pi_i = e - t - g - b + \frac{m}{n}[(n-2)g + g_K + e - t] = (n-1)b + \frac{m}{n}[(n-2)g + g_K + e - t]$$

If Punish:
($b_K = 0$)

$$\pi_K = e - t - g_K + \frac{m}{n}[(n-2)g + g_K + e - t] - rd_{AK}$$

$$\pi_A = (n-2)b + \frac{m}{n}[(n-2)g + g_K + e - t]$$

$$\pi_A = \pi_K \Rightarrow e - t - g_K - rd_{AK} = (n-2)b \Rightarrow d_{AK} = \frac{2b + (g - g_K)}{r}$$

$$u_{A,P} = (n-2)b + \frac{m}{n}[(n-2)g + g_K + e - t] - \alpha_A \frac{n-2}{n-1} b$$

If Accept Bribe:
$(d_{AK} = 0)$

$$\pi_K = e - t - g_K - b_K + \frac{m}{n}[(n-2)g + g_K + e - t]$$

$$\pi_A = (n-2)b + b_K + \frac{m}{n}[(n-2)g + g_K + e - t]$$

$$u_{A,B} = (n-2)b + b_K + \frac{m}{n}[(n-2)g + g_K + e - t] - \alpha_A \frac{n(b - b_K) + (g - g_K)}{n-1}$$

Punish if:

$$(n-2)b - \alpha_A \frac{n-2}{n-1} b > (n-2)b + b_K - \alpha_A \frac{n(b - b_K) + (g - g_K)}{n-1} \Rightarrow 2\alpha_A b > (\alpha_A n + n - 1) b_K - \alpha_A (g - g_K)$$

Case 5. $g_K > g$ and $b_K < b$:

$$\pi_K = e - t - g_K - b_K + \frac{m}{n}[(n-2)g + g_K + e - t] - r d_{AK}$$

$$\pi_A = (n-2)b + b_K + \frac{m}{n}[(n-2)g + g_K + e - t]$$

$$\pi_i = e - t - g - b + \frac{m}{n}[(n-2)g + g_K + e - t] = (n-1)b + \frac{m}{n}[(n-2)g + g_K + e - t]$$

If Punish:
$(b_K = 0)$

$$\pi_K = e - t - g_K + \frac{m}{n}[(n-2)g + g_K + e - t] - r d_{AK}$$

$$\pi_A = (n-2)b + \frac{m}{n}[(n-2)g + g_K + e - t]$$

$$\pi_A = \pi_K \Rightarrow e - t - g_K - r d_{AK} = (n-2)b \Rightarrow d_{AK} = \frac{2b + (g - g_K)}{r}$$

$$u_{A,P} = (n-2)b + \frac{m}{n}[(n-2)g + g_K + e - t] - \alpha_A \frac{n-2}{n-1} b$$

If Accept Bribe:
$(d_{AK} = 0)$

$$\pi_K = e - t - g_K - b_K + \frac{m}{n}[(n-2)g + g_K + e - t]$$

$$\pi_A = (n-2)b + b_K + \frac{m}{n}[(n-2)g + g_K + e - t]$$

$$u_{A,B} = (n-2)b + b_K + \frac{m}{n}[(n-2)g + g_K + e - t] - \alpha_A \frac{n(b-b_K)+(g-g_K)}{n-1}$$

Punish if:

$$(n-2)b - \alpha_A \frac{n-2}{n-1} b > (n-2)b + b_K - \alpha_A \frac{n(b-b_K)+(g-g_K)}{n-1} \Rightarrow 2\alpha_A b > (\alpha_A n + n - 1) b_K - \alpha_A (g - g_K)$$