Moral Rules and Social Preferences in Cooperation problems

BY ERNESTO M. GAVASSA-PÉREZ*

Online Appendix A: Proofs

Table of Contents

A.1. Fixing some notation

The public goods game we consider is a 2-player, one-shot game. The relevant data from the P-experiment's strategy method (i.e., the conditional contribution task) is sequential in nature. To fix some notation before proceeding, we will henceforth refer to the two players in a group as player i and player j. We fix subject i's optimal contribution schedule in the conditional contribution task to be referred to as c_i^*; which will involve an optimal contribution against each potential contribution of the other player (that is, against each g_j). To make the notation more salient, and less prone to confusion with letter c, which we already use to denote the optimal contribution schedule, we opt to call a given contribution by player i as g_i, and a given contribution of player j by g_j. In mathematical terms, g_i and g_j are but generic contributions feasible for each player and lie within the sets $g_i \in A_i \coloneqq \{0,10,20,30\}$, and $g_j \in A_j \coloneqq \{0,10,20,30\}$. Hence, the cartesian product $A_i \times A_j$ refers to the set containing all strategy combinations of players i and j, and we denominate $\langle g_i, g_j \rangle$ (or, for notational compactness, g_i, g_j when within a parenthesis) to refer to a generic strategy combination of i and j that lie within the cartesian product defined earlier. The material payoff of player i (and analogously for player j) is represented by the following function:

$$\pi_i(g_i, g_j) = 30 - g_i + m \times (g_i + g_j)$$

Where $m \in \left(\frac{1}{n}, 1\right)$ for a social dilemma and $m \in (1, \infty)$ for a common interest game. At some points we will refer to $\underline{m}$ as an arbitrarily small value of the marginal per capita return and to $\overline{m}$ as an arbitrarily large value of the marginal per capita return to the public good. In all such instances, $\underline{m}$ will refer to a social dilemma game (that is, $\underline{m} \in \left(\frac{1}{n}, 1\right)$) and $\overline{m}$ will refer to a common interest game (that is, $\overline{m} \in (1, \infty)$).

A.2. Predictions of theories regarding contribution preferences

A.2.1.. An important lemma

For all the proofs that follow, and to shorten the derivations, we will use extensively a result. We summarise such a result in the following lemma:

Lemma 0. *In the aforementioned two-player, one-shot, public goods game, with the payoff functions $\pi_i(g_i, g_j)$ and $\pi_j(g_i, g_j)$ denoting, respectively, the payoffs of player i and player j from the strategy combination $\langle g_i, g_j \rangle \in A_i \times A_j$, it follows that:*

(a) $\pi_i(g_i, g_j) > \pi_j(g_i, g_j)\ iff\ g_i < g_j$.

(b) $\pi_i(g_i, g_j) - \pi_j(g_i, g_j) = g_j - g_i$ *and* $\pi_j(g_i, g_j) - \pi_i(g_i, g_j) = g_i - g_j$

Proof.

First part of the proof: Proving lemma 0 (a)

Let's consider player i makes an arbitrarily small contribution $\underline{g_i}$, and let further $g_j > \underline{g_i}$ be the case. Then, the material payoff of player i when contributing $\underline{g_i}$, given that the other player contributes g_j is given by:

$$\pi_i\left(\underline{g_i}, g_j\right) = 30 - \underline{g_i} + m \times \left(\underline{g_i} + g_j\right)$$

And the payoff of player j given $\underline{g_i}$ and g_j is equivalent to the following expression:

$$\pi_j\left(\underline{g_i}, g_j\right) = 30 - g_j + m \times \left(\underline{g_i} + g_j\right)$$

Subtracting the latter from the former, we get:

$$\pi_i\left(\underline{g_i}, g_j\right) - \pi_j\left(\underline{g_i}, g_j\right) = 30 - \underline{g_i} + m \times \left(\underline{g_i} + g_j\right) - \left\{30 - g_j + m \times \left(\underline{g_i} + g_j\right)\right\}$$

Expanding the curly brackets, we get:

$$\pi_i\left(\underline{g_i}, g_j\right) - \pi_j\left(\underline{g_i}, g_j\right) = 30 - \underline{g_i} + m \times \left(\underline{g_i} + g_j\right) - 30 + g_j - m \times \left(\underline{g_i} + g_j\right)$$

Simplifying, we get:

$$\pi_i\left(\underline{g_i}, g_j\right) - \pi_j\left(\underline{g_i}, g_j\right) = g_j - \underline{g_i}$$

Given that $\underline{g_i} < g_j$, it then follows that $g_j - \underline{g_i} > 0$. Hence,

$$\pi_i\left(\underline{g_i}, g_j\right) - \pi_j\left(\underline{g_i}, g_j\right) > 0$$

Bringing $\pi_j\left(\underline{g_i}, g_j\right)$ to the RHS, we get:

$$\pi_i\left(\underline{g_i}, g_j\right) > \pi_j\left(\underline{g_i}, g_j\right)$$

Which proves lemma 0 (a).

<u>Second part of the proof: Proving lemma 0 (b)</u>

Now, substituting $\underline{g_i}$ by g_i in the derivations above it is straightforward to see that

$$\pi_i(g_i, g_j) - \pi_j(g_i, g_j) = g_j - g_i$$

Additionally, multiplying both hand sides by -1 we can see that:

$$\pi_j(g_i, g_j) - \pi_i(g_i, g_j) = g_i - g_j$$

Which proves lemma 0 (b).

QED.

A.2.2. Homo Economicus preferences – Proof of proposition 1

Proposition 1. *If subject i maximizes the utility function $U_i^{HE}(g_i, g_j) = \pi_i(g_i, g_j)$, where $\pi_i(g_i, g_j)$ denotes the material payoff of person i for the strategy combination in which i contributes g_i and the other player g_j, subject i's optimal contributions will be $c_i^* = g_i = 0\ \forall g_j \in A_j$ (resp. $c_i^* = g_i = 30\ \forall g_j \in A_j$) in the SDG (resp. CIG).*

Proof.

To see this, note that $\frac{\partial U_i^{HE}(g_i,g_j)}{\partial g_i} = m - 1$, which is negative for any social dilemma (as $m < 1$) and positive for any CIG (as $m > 1$). Therefore, it follows that $c_i^* = g_i = 0 \; \forall g_j \in A_j$ ($c_i^* = g_i = 30 \; \forall g_j \in A_j$) is the solution to subject i's maximization problem in the SDG (CIG).

QED.

A.2.3. Inequality Aversion Preferences

A.2.3.1. Proof of proposition 2

Proposition 2. *If subject* i *maximizes the utility function* $U_i^{FS}\left(\pi_i(g_i, g_j), \pi_j(g_i, g_j)\right)$*, where* i *contributes* g_i *and the other player contributes* g_j*, then subject* i*'s contribution attitudes, denoted as* c_i^**, will be*

(i), in the Social Dilemma,

$$c_i^* = \begin{cases} g_i = 0 \; \forall g_j \in A_j & iff \; \beta_i < 1 - \underline{m} \\ g_i = g_j \forall g_j \in A_j & iff \; \beta_i > 1 - \underline{m} \\ g_i \in [0, g_j] \; \forall g_j \in A_j & iff \; \beta_i = 1 - \underline{m} \end{cases}$$

(ii), in the Common Interest Game,

$$c_i^* = \begin{cases} g_i = 30 \; \forall g_j \in A_j & iff \; \alpha_i < \overline{m} - 1 \\ g_i = g_j \; \forall \; g_j \in A_j & iff \; \alpha_i > \overline{m} - 1 \\ g_i \in [g_j, 30] \; \forall g_j \in A_j & iff \; \alpha_i = \overline{m} - 1 \end{cases}$$

Proof.

<u>First part of the proof: proving (i)</u>

Step 1: Recall necessary functions.

First, let's recall the utility function we use to measure inequality aversion preferences:

$$U_i^{\mathrm{FS}}(\pi_i, \pi_\mathrm{j}) \coloneqq \pi_i - \alpha_i * Max\{\pi_\mathrm{j} - \pi_i, 0\} - \beta_i * Max\{\pi_i - \pi_\mathrm{j}, 0\}$$

Step 2: Calculate the utility function of player i for cases where $g_i < g_j$.

Let's assume that player i contributes less than player j. To keep the notation consistent throughout the text, let's denote such a contribution as $\underline{g_i}$. Then, the utility function of a Fehr-Schmidt player i will take the following form:

$$U_i^{FS}\left(\pi_i\left(\underline{g_i}, g_j\right), \pi_j\left(\underline{g_i}, g_j\right)\right) = \pi_i\left(\underline{g_i}, g_j\right) - \beta_i \times \left(\pi_i\left(\underline{g_i}, g_j\right) - \pi_j\left(\underline{g_i}, g_j\right)\right)$$

Substituting $\pi_i\left(\underline{g_i}, g_j\right) = 30 - \underline{g_i} + m \times \left(\underline{g_i} + g_j\right)$ in the first term of the RHS and using the results of lemma 0 (b) above to simplify the last term of the RHS, $U_i^{FS}(\pi_i, \pi_j)$ collapses to:

$$U_i^{FS}\left(\pi_i\left(\underline{g_i}, g_j\right), \pi_j\left(\underline{g_i}, g_j\right)\right) = 30 - \underline{g_i} + m \times \left(\underline{g_i} + g_j\right) - \beta_i \times \left(g_j - \underline{g_i}\right)$$

Step 3: Calculate the utility function of player i for cases where $g_i > g_j$.

Let's now consider the case where player i contributes more than player j, and let's denominate such a contribution as $\bar{g}_i > g_j$. Analogously to the previous step, substituting $\pi_i(\bar{g}_i, g_j) = 30 - \bar{g}_i + m \times (\bar{g}_i + g_j)$ in the first term of the RHS and using, again, the results from lemma 0 (b), we can rewrite the utility function as follows:

$$U_i^{FS}\left(\pi_i(\bar{g}_i, g_j), \pi_j(\bar{g}_i, g_j)\right) = 30 - \bar{g}_i + m \times (\bar{g}_i + g_j) - \alpha_i \times (\bar{g}_i - g_j)$$

Step 4: Write the utility function of player i for cases where $g_i = g_j$.

By lemma 0 (b), we know that $\pi_j(g_i, g_j) - \pi_i(g_i, g_j) = g_i - g_j$. Hence, whenever $g_i = g_j$, then $\pi_j(g_i, g_j) - \pi_i(g_i, g_j) = 0$. Substituting this into our utility function, we get:

$$U_i^{FS}\left(\pi_i(g_i, g_j), \pi_j(g_i, g_j)\right) = \pi_i(g_i, g_j) - \beta_i \times (0)$$

And, hence, $U_i^{FS}\left(\pi_i(g_i, g_j), \pi_j(g_i, g_j)\right) = U_i^{HE}(g_i, g_j)\ \forall g_i = g_j$.

Step 5: Write the utility function of player i for all possible cases of $g_i \gtreqless g_{-i}$.

Given the results of steps 2 to 4, we can then write the Fehr-Schmidt utility function more compactly as:

$$U_i^{FS}\left(\pi_i(g_i, g_j), \pi_j(g_i, g_j)\right) = \begin{cases} 30 - g_i + m \times (g_i + g_j) - \beta_i \times (g_j - g_i)\ if\ g_i < g_j \\ 30 - g_i + m \times (g_i + g_j)\ if\ g_i = g_j \\ 30 - g_i + m \times (g_i + g_j) - \alpha_i \times (g_i - g_j)\ if\ g_i > g_j \end{cases}$$

Step 6: Finding person i's first derivative with respect to g_i.

Taking the first derivative of the linear utility function with respect to g_i, we get

$$\frac{\partial U_i^{FS}\left(\pi_i(g_i, g_j), \pi_j(g_i, g_j)\right)}{\partial g_i} = \begin{cases} -1 + m + \beta_i\ if\ g_i < g_j \\ -1 + m\ if\ g_i = g_j \\ -1 + m - \alpha_i\ if\ g_i > g_j \end{cases}$$

Step 7: Impose in the previous derivative $m = \underline{m} < 1$.

Thus, for a generic value $\underline{m} \in \left(\frac{1}{n}, 1\right)$, the previous first derivative reads:

$$\frac{\partial U_i^{FS}\left(\pi_i(g_i, g_j), \pi_j(g_i, g_j)\right)}{\partial g_i} = \begin{cases} -1 + \underline{m} + \beta_i\ if\ g_i < g_j \\ -1 + \underline{m}\ if\ g_i = g_j \\ -1 + \underline{m} - \alpha_i\ if\ g_i > g_j \end{cases}$$

Step 8: Prove that $c_i^* = g_i > g_j$ *is not optimal given all the potential values of* α_i and $\underline{m}$.

As $\alpha_i \geq 0$ and $\underline{m} < 1$, then from the last step it follows that, if $g_i > g_j$, then

$$\frac{\partial U_i^{FS}\big(\pi_i(g_i,g_j),\pi_j(g_i,g_j)\big)}{\partial g_i} = -1 + \underline{m} - \alpha_i = -1 + (< 1) - (\geq 0) = (< 0) + (\leq 0) = (< 0).$$

It follows that the marginal utility will always be strictly negative for $g_i > g_j$, and, given the linearity of the utility function, person i's optimal contribution against g_j will never lie within the range defined by $g_i > g_j$.

Step 9: Give the range of values of β_i for which the marginal utility is positive (resp. negative; resp. zero), given $g_i < g_j$.

Turning to the case where $g_i < g_j$, we have three different outcomes:

When $g_i < g_j$, then

- $\frac{\partial U_i^{FS}}{\partial g_i} < 0 \; iff \; \beta_i < 1 - \underline{m}$
- $\frac{\partial U_i^{FS}}{\partial g_i} > 0 \; iff \; \beta_i > 1 - \underline{m}$
- $\frac{\partial U_i^{FS}}{\partial g_i} = 0 \; iff \; \beta_i = 1 - m$

Step 10: Outline c_i^* for an SDG (i.e., given $\underline{m}$) in lieu of the previous steps.

Given steps 8 and 9, and the linearity of U_i^{FS}, i's best response against each potential g_j (that is, c_i^*) in the SDG will be given by:

$$c_i^* = \begin{cases} g_i = 0 \; \forall g_j \in A_j & if \; \beta_i < 1 - \underline{m} \\ g_i \in [0, g_j] \; \forall g_j \in A_j & if \; \beta_i = 1 - \underline{m} \\ g_i = g_j \; \forall \; g_j \in A_j & if \; \beta_i > 1 - \underline{m} \end{cases}$$

This follows from three facts:

1. First, note that whenever $\beta_i < 1 - \underline{m}$, then $\left(\forall \langle g_i g_j \rangle \in A_i \times A_j\right), \frac{\partial U_i^{FS}\left(\pi_i(g_i,g_j),\pi_j(g_i,g_j)\right)}{\partial g_i} < 0$. Hence, $g_i = 0 \; \forall g_j \in A_j$ will maximise person i's contribution against each possible g_j.
2. Second, note that, whenever $\beta_i = 1 - \underline{m}$, then $\left(\forall \langle g_i g_j \rangle \in A_i \times A_j\right), \frac{\partial U_i^{FS}\left(\pi_i(g_i,g_j),\pi_j(g_i,g_j)\right)}{\partial g_i} = 0 \; iff \, g_i \in [0, g_j]$; implying that person i's utility for all $g_i \leq g_j$ will be the same; all being optimal contributions.
3. Third, note that, whenever $\beta_i < 1 - \underline{m}$, then $\left(\forall \langle g_i g_j \rangle \in A_i \times A_j\right), \frac{\partial U_i^{FS}\left(\pi_i(g_i,g_j),\pi_j(g_i,g_j)\right)}{\partial g_i} > 0 \; iff \; g_i < g_j$ and $\frac{\partial U_i^{FS}\left(\pi_i(g_i,g_j),\pi_j(g_i,g_j)\right)}{\partial g_i} < 0 \; iff \; g_i \geq g_j$. Hence, person i's utility will be maximised, in such cases, at $g_i = g_j$.

<u>*Second part of the proof: proving (ii)*</u>

Step 11: Impose in the derivative $m = \overline{m} > 1$.

For a generic value $\overline{m}$, the previous first derivative is equivalent to:

$$\frac{\partial U_i^{FS}\left(\pi_i(g_i,g_j),\pi_j(g_i,g_j)\right)}{\partial g_i} = \begin{cases} -1 + \overline{m} + \beta_i \; if \; g_i < g_j \\ -1 + \overline{m} \; if \; g_i = g_j \\ -1 + \overline{m} - \alpha_i \; if \; g_i > g_j \end{cases}$$

Step 12: Prove that $g_i < g_j$ is not optimal given all the potential values of β_i and $\overline{m}$.

As $\beta_i \geq 0$ and $\overline{m} > 1$, then from the derivate it follows that, if $g_i < g_j$, then

$$\frac{\partial U_i^{FS}\left(\pi_i(g_i,g_j),\pi_j(g_i,g_j)\right)}{\partial g_i} = -1 + \overline{m} + \beta_i = -1 + (> 1) + (\geq 0) = (> 0) + (\geq 0)$$
$$= (> 0)$$

It follows that the marginal utility will always be strictly positive for $g_i < g_j$; and, given the linearity of the utility function, person i's optimal contribution will never lie within the range defined by $g_i < g_j$.

Step 13: Give the range of values of α_i for which the marginal utility is positive (resp. negative; resp. zero) given $g_i > g_j$.

Turning to the case where $g_i > g_j$, we have three different outcomes:

- $\frac{\partial U_i^{FS}}{\partial g_i} < 0 \; iff \; \alpha_i > \overline{m} - 1$

- $\frac{\partial U_i^{FS}}{\partial g_i} > 0 \; iff \; \alpha_i < \overline{m} - 1$

- $\frac{\partial U_i^{FS}}{\partial g_i} = 0 \; iff \; \alpha_i = \overline{m} - 1$

Step 14: Outline c_i^* for a CIG (i.e., given $\overline{m}$) in lieu of the previous steps

Given steps 12 and 13, and the linearity of U_i^{FS}, i's best response against g_j (that is, c_i^*) in the CIG will be given by:

$$c_i^* = \begin{cases} g_i = 30 \; \forall g_j \in A_j & iff \; \alpha_i < \overline{m} - 1 \\ g_i \in [g_j, 30] \; \forall g_j \in A_j & iff \; \alpha_i = \overline{m} - 1 \\ g_i = g_j \; \forall g_j \in A_j & iff \; \alpha_i > \overline{m} - 1 \end{cases}$$

This follows from three facts:

1. First, note that whenever $\alpha_i < \overline{m} - 1$, then $\left(\forall \langle g_i g_j \rangle \in A_i \times A_j\right), \frac{\partial U_i^{FS}\left(\pi_i(g_i, g_j), \pi_j(g_i, g_j)\right)}{\partial g_i} > 0$. Hence, $c_i^* = g_i = 30 \; \forall g_j \in A_j$ will maximise person i's contribution against each possible g_j.

2. Second, note that, whenever $\alpha_i = \overline{m} - 1$, then $\left(\forall \langle g_i g_j \rangle \in A_i \times A_j\right), \frac{\partial U_i^{FS}\left(\pi_i(g_i, g_j), \pi_j(g_i, g_j)\right)}{\partial g_i} = 0 \; iff \; g_i \in \left[g_j, 30\right]$, implying that person i's utility for all $g_i \geq g_j$ will be the same, all being optimal contributions.
3. Third, note that, whenever $\alpha_i > \overline{m} - 1$, then $\left(\forall \langle g_i g_j \rangle \in A_i \times A_j\right), \frac{\partial U_i^{FS}\left(\pi_i(g_i, g_j), \pi_j(g_i, g_j)\right)}{\partial g_i} < 0 \; iff \; g_i > g_j$ and $\frac{\partial U_i^{FS}\left(\pi_i(g_i, g_j), \pi_j(g_i, g_j)\right)}{\partial g_i} > 0 \; iff \; g_i < g_j$. Hence, person i's utility will be maximised, in such cases, at $g_i = g_j$.

QED.

A.2.3.2. Other results involving inequality aversion preferences

We use the results from proposition 2 to provide, in corollary 2.1, the precise contribution attitudes in the SDG and CIG that we use in chapter 4. Additionally, we provide another main result besides proposition 2. Namely, that for some joint values of $\underline{m}$ and $\overline{m}$ person i cannot be a perfect conditional cooperator (i.e., $g_i = g_j \; \forall \; g_j \in A_j$) in the SDG and an unconditional cooperator in the CIG (i.e., $g_i = 30 \; \forall g_j \in A_j$), as it would require a violation of the parameter restrictions of Fehr-Schmidt (i.e., it would require $\beta_i > \alpha_i$). Hence, inequality aversion cannot predict perfect conditional cooperation in the SDG and unconditional cooperation in the CIG. We summarise this second result in corollary2.2. Additionally, corollary 2.3 shows that, for the values of $\underline{m}$ and $\overline{m}$ used in the experiments of chapter 4, the inequality aversion model cannot predict conditional co-operation in the SDG and unconditional co-operation in the CIG.

Corollary 2.1. *If subject i maximizes the utility function* $U_i^{FS}\left(\pi_i(g_i, g_j), \pi_j(g_i, g_j)\right)$*, and* $\underline{m} = 0.6$ in the SDG and $\overline{m} = 1.2$ in the CIG, then

(a) has $\beta_i < 0.4$ *(resp.* $\beta_i = 0.4$*; resp.* $\beta_i > 0.4$*), then subject i's cooperation attitude in the SDG will be* $c_i^* = g_i = 0 \; \forall g_j \in A_j$ *(resp.* $c_i^* = g_i \in \left[0, g_j\right] \; \forall g_j \in A_j$*; resp.* $c_i^* = g_i = g_j \; \forall \; g_j \in A_j$*).*

(b) has $\alpha_i < 0.2$ *(resp.* $\alpha_i = 0.2$*; resp.* $\alpha_i > 0.2$*), then subject i's cooperation attitude in the CIG will be* $c_i^* = g_i = 30 \; \forall g_j \in A_j$ *(resp.* $c_i^* = g_i \in \left[g_j, 30\right] \; \forall g_j \in A_j$*; resp.* $c_i^* = g_i = g_j \; \forall \; g_j \in A_j$*).*

Proof.

Substituting $\underline{m} = 0.6$ and $\overline{m} = 1.2$ in the cooperation attitudes found in proposition 2, we get the two following expressions:

$$c_i^* = \begin{cases} g_i = 0 \ \forall g_j \in A_j & if\ \beta_i < 1 - 0.6 \\ g_i \in [0, g_j] \ \forall g_j \in A_j & if\ \beta_i = 1 - 0.6 \\ g_i = g_j \ \forall\ g_j \in A_j & if\ \beta_i > 1 - 0.6 \end{cases}$$

$$c_i^* = \begin{cases} g_i = 30 \ \forall g_j \in A_j & iff\ \alpha_i < 1.2 - 1 \\ g_i \in [g_j, 30] \ \forall g_j \in A_j & iff\ \alpha_i = 1.2 - 1 \\ g_i = g_j \ \forall g_j \in A_j & iff\ \alpha_i > 1.2 - 1 \end{cases}$$

Which, after simplifying, become:

$$c_i^* = \begin{cases} g_i = 0 \ \forall g_j \in A_j & if\ \beta_i < 0.4 \\ g_i \in [0, g_j] \ \forall g_j \in A_j & if\ \beta_i = 0.4 \\ g_i = g_j \ \forall\ g_j \in A_j & if\ \beta_i > 0.4 \end{cases}$$

$$c_i^* = \begin{cases} g_i = 30 \ \forall g_j \in A_j & iff\ \alpha_i < 0.2 \\ g_i \in [g_j, 30] \ \forall g_j \in A_j & iff\ \alpha_i = 0.2 \\ g_i = g_j \ \forall g_j \in A_j & iff\ \alpha_i > 0.2 \end{cases}$$

QED.

Corollary 2.2. *If subject* i *maximizes the utility function* $U_i^{FS}\left(\pi_i(g_i, g_j), \pi_j(g_i, g_j)\right)$, *where* i *contributes* g_i *and the other player contributes* g_j, *and further* $2 > \underline{m} + \overline{m}$ *holds true, then subject* i *will be a perfect conditional co-operator in the SD and an unconditional co-operator in the CIG iff* $\beta_i > \alpha_i$.

Proof.

Step 1: Provide the conditions for perfect conditional cooperation in the SDG and unconditional cooperation in the CIG.

Given proposition 2, Subject i will only be a perfect conditional cooperator (i.e., $g_i = g_j \; \forall \; g_j \in A_j$) in the SDG iff the following condition holds:

$$\beta_i > 1 - \underline{m}$$

Additionally, given proposition 2, Subject i will only be an unconditional cooperator (i.e., $g_i = 30 \; \forall \; g_j \in A_j$) in the CIG iff the following condition holds:

$$\alpha_i < \overline{m} - 1$$

Step 2: Establish the result by contradiction.

Assume $2 > \underline{m} + \overline{m}$, that subject i is a perfect conditional co-operator in the SD and an unconditional co-operator in the CIG, and that $\alpha_i > \beta_i$ holds true at the same time. Then, by using the two previous conditions and imposing $\alpha_i > \beta_i$, we would get:

$$\overline{m} - 1 > \alpha_i > \beta_i > 1 - \underline{m}$$

From which it trivially follows that:

$$\overline{m} - 1 > 1 - \underline{m}$$

And, hence,

$$\overline{m} + \underline{m} > 2$$

Thus, if $2 > \underline{m} + \overline{m}$, subject i is a perfect conditional co-operator in the SD and an unconditional co-operator in the CIG, and $\alpha_i > \beta_i$ hold true at the same time, it must be that $2 > \underline{m} + \overline{m}$ and $2 < \underline{m} + \overline{m}$ hold true at the same time, which is a contradiction. Therefore, if subject i is a perfect conditional co-operator in the SD and an unconditional co-operator in the CIG, and it happens to be that $2 > \underline{m} + \overline{m}$, then $\alpha_i < \beta_i$ must be true.

QED.

A.2.4. Reciprocity preferences

A.2.4.1. Fixing some notation specific to sequential reciprocity

In the next pages we present the theoretical derivations for the reciprocity model of Dufwenberg and Kirchsteiger (2004). From now on, we use $(p', g_i = x; q', g_i \neq x)$ as a notation to describe the probabilities associated with contribution levels $g_i = x$ and $g_i \neq x$, which represent nothing but the first order beliefs. Hence, we use $(p', g_i = 0; q', g_i = 10; r', g_i = 20; 1 - p' - q' - r', g_i = 30)$ to refer to the probabilities associated to each of the possible contribution levels in our games. We denote the probabilities associated with second order beliefs as p'', q'', and so on. Additionally, in the contribution table task we assume that the contribution of the other person in each cell represents the first order belief with certainty of the responder. This is the case as, given the comment in Fischbacher et al (2001), the responses to each cell in the strategy method, given the incentive compatible mechanism used, can be seen as the responses of a second mover to each potential move of the first mover. And, given the belief updating mechanism in Dufwenberg and Kirchsteiger (2004), at each node the second mover updates his belief to reflect what has been played by the first mover, hence collapsing the first order belief to the strategy that led to the node being played.

As a reminder, below is the utility function of person i if person i were to follow Dufwenberg and Kirchsteiger's (2004) model of reciprocity:

$$U_i^{DK}(\pi_i, \pi_j) = \pi_i\left(g_i(h), b_{ij}(h), c_{iji}(h)\right)$$
$$= \pi_i\left(g_i(h), b_{ij}(h)\right) + Y_{i,j} \times \kappa_{ij}\left(g_i(h), b_{ij}(h)\right) \times \lambda_{iji}\left(b_{ij}(h), c_{iji}(h)\right)$$

Where Y_{ij} is a parameter measuring the strength of reciprocal motivations, $\kappa_{ij}\left(g_{ij}(h), b_{ij}(h)\right)$ is a function measuring how kind is person i being with person j, $\lambda_{ij}\left(b_{ij}(h), c_{iji}(h)\right)$ is a function measuring how kind person i perceives person j is being towards him and $g_i(h)$, $b_{ij}(h)$ and $c_{ij}(h)$ are, respectively, the contribution, first- and second-order beliefs of person i at node h. Given that person i is a second mover, $b_{ij}(h)$ is updated to reflect the contribution level of the first mover, person j; being, hence, possible an alternative notation $b_{ij}(h) = g_j$.

A.2.4.2. Proof of proposition 3

Proposition 3. *If subject* i *maximizes the utility function* $U_i^{DK}\left(\pi_i(g_i, g_j), \pi_j(g_i, g_j)\right)$, *where* i *contributes* g_i, *the other player contributes* g_j, *and the other player moves first and subject* i *second, and where we denote* c_i^* *as subject* i*'s optimal contribution, then subject* i *will*

(i), in the Social Dilemma,

- *(a) do* $c_i^* = g_i = 0$ *against* $g_j \in \{0,10\}$ *regardless of* $Y_{i,j}$
- *(b) do* $c_i^* = g_i = 0$ *against* $g_j \in \{20,30\}$ *iff* $Y_{i,j} < \frac{1-\underline{m}}{\underline{m}^2 \times (g_j - 15)}$
- *(c) do* $c_i^* = g_i \in A_i$ *against* $g_j \in \{20,30\}$ *iff* $Y_{i,j} = \frac{1-\underline{m}}{\underline{m}^2 \times (g_j - 15)}$
- *(d) do* $c_i^* = g_i = 30$ *against* $g_j \in \{20,30\}$ *iff* $Y_{i,j} > \frac{1-\underline{m}}{\underline{m}^2 \times (g_j - 15)}$

(ii), in the Common Interest Game,

- *(e) do* $c_i^* = g_i = 30$ *against* $g_j = 30$ *regardless of* $Y_{i,j}$
- *(f) do* $c_i^* = g_i = 0$ *against* $g_j \in \{0,10,20\}$*) iff* $Y_{i,j} > \frac{\overline{m}-1}{\overline{m}^2 \times (30 - g_j)}$
- *(g) do* $c_i^* = g_i \in A_i$ *against* $g_j \in \{20,30\}$ *iff* $Y_{i,j} = \frac{\overline{m}-1}{\overline{m}^2 \times (30 - g_j)}$
- *(h) do* $c_i^* = g_i = 30$ *against* $g_j \in \{0,10,20\}$*) iff* $Y_{i,j} < \frac{\overline{m}-1}{\overline{m}^2 \times (30 - g_j)}$

Proof.

The proof for this proposition is very long, so we start by summarising the approach we take before the reader engages with the reading of the proof. The first steps will involve computing the kindness and perceived kindness functions of person i for a generic level of the other person. The next steps will involve substituting those functional forms into $U_i^{DK}\left(\pi_i(g_i, g_j), \pi_j(g_i, g_{ji})\right)$ to get the utility function of person i in terms, only, of g_i and g_j. We, then, compute the first order derivative of $U_i^{DK}\left(\pi_i(g_i, g_j), \pi_j(g_i, g_{ji})\right)$ with respect to g_i to find the optimal contribution levels of g_i. This is done, as was the case with inequality aversion preferences, by assessing if the utility function is either increasing or decreasing in g_i at every level of g_j. We will carry out this process separately for the SDG and the CIG as the set of efficient strategies is different for both games, making the functional form of the kindness and perceived kindness functions to differ across games.

Step 1: find the kindness function (κ_{ij}) of subject i in the SDG.

At generic contribution levels g_j and g_i, we can write the kindness function as:

$$\kappa_{ij}\left(g_{ij}(h), b_{ij}(h)\right) = \pi_j\left(g_i, b_{ij}(h)\right) - \frac{\max \pi_j\left(g_i, b_{ij}(h)\right) + \min \pi_j\left(g_i, b_{ij}(h)\right)}{2}$$

Given that $\pi_j(g_i, g_j) = 30 - g_j + \underline{m} \times (g_i + g_j)$, taking the first derivative with respect to g_i, we get:

$$\frac{\partial \pi_j(g_i, g_j)}{\partial g_i} = \underline{m} > 0$$

Hence, the payoff of person j is increasing in g_i. This means that the payoff of person j will be maximised, given g_j, at the highest contribution level of person i and will be minimised at the lowest contribution level of person i. Those are, respectively, $g_i = 30$ and $g_i = 0$. Additionally, and given that person j is the first mover, then $b_{ij}(h) = g_j$. Hence, we can rewrite the kindness function as:

$$\kappa_{ij}\left(g_{ij}(h), b_{ij}(h) = g_j\right) = \pi_j\left(g_i, g_j\right) - \frac{\pi_j\left(g_i = 30, g_j\right) + \pi_j\left(g_i = 0, g_j\right)}{2}$$

Substituting $\pi_j\left(g_i, g_j\right)$ by the material payoff function outlined above, and g_i by 0 and 30 where appropriate, we get:

$$\begin{aligned}\kappa_{ij}&\left(g_{ij}(h), g_j\right) \\ &= 30 - g_j + \underline{m} \times \left(g_i + g_j\right) \\ &\quad - \frac{30 - g_j + \underline{m} \times \left(30 + g_j\right) + 30 - g_j + \underline{m} \times \left(g_j\right)}{2}\end{aligned}$$

Grouping the terms in the numerator, and taking $\underline{m}$ as a common factor in the numerator, we get:

$$\kappa_{ij}\left(g_{ij}(h), g_j\right) = 30 - g_j + \underline{m} \times \left(g_i + g_j\right) - \frac{60 - 2 \times g_j + \underline{m} \times \left(30 + 2 \times g_j\right)}{2}$$

Which can be rewritten as:

$$\kappa_{ij}\left(g_{ij}(h), g_j\right) = 30 - g_j + \underline{m} \times \left(g_i + g_j\right) - \left(30 - g_j + \underline{m} \times \left(15 + g_j\right)\right)$$

Expanding the expression $-\left(30 - g_j + \underline{m} \times \left(15 + g_j\right)\right)$, we get:

$$\kappa_{ij}\left(g_{ij}(h), g_j\right) = 30 - g_j + \underline{m} \times \left(g_i + g_j\right) - 30 + g_j - \underline{m} \times \left(15 + g_j\right)$$

Simplifying, we get:

$$\kappa_{ij}\left(g_{ij}(h), g_j\right) = \underline{m} \times \left(g_i + g_j\right) - \underline{m} \times \left(15 + g_j\right)$$

Using $\underline{m}$ as a common factor, we get:

$$\kappa_{ij}\left(g_{ij}(h), g_j\right) = \underline{m} \times \left(g_i + g_j - 15 - g_j\right)$$

And, finally, simplifying we get:

$$\kappa_{ij}\left(g_{ij}(h), g_j\right) = \underline{m} \times (g_i - 15)$$

Step 2: find the perceived kindness function (λ_{iji}) of subject i in the SDG.

To compute the perceived kindness function, let us denominate $c_{iji}(h) = (p'', g_i = 0; q'', g_i = 10; r'', g_i = 20; 1 - p'' - q'' - r'', g_i = 30)$ as the probability distribution of the second-order belief of player i. Unlike the first-order belief, the first mover did not know what player 2 was going to do when he or she decided to contribute g_j. Hence, we assume that the second mover believes that the first mover didn't know what the second mover was going to do when first mover chose g_j. The probability distribution $c_{iji}(h)$ over the second-order belief captures that uncertainty. We use such generic probability distribution to denote the belief that player i has about the belief of player *j* of player i's contribution when player *j* was making the decision of contributing g_j (contribution at the initial node). For compactness in the notation, we just write $c_{iji}(h)$ instead of writing $c_{iji}(h) = (p'', g_i = 0; q'', g_i = 10; r'', g_i = 20; 1 - p'' - q'' - r'', g_i = 30)$ in our definition of the perceived kindness function of person i. We can define the perceived kindness function of player i as:

$$\lambda_{iji}\left(b_{ij}(h), c_{iji}(h)\right) = \pi_i\left(b_{ij}(h), c_{iji}(h)\right) - \frac{\max \pi_i\left(b_{ij}(h), c_{iji}(h)\right) + \min \pi_i\left(b_{ij}(h), c_{iji}(h)\right)}{2}$$

As noted before, the payoff function of a given player is increasing in the contribution of the other player. Hence, person i's payoff will be maximised at $b_{ij}(h) = 30$ and minimised at $b_{ij}(h) = 0$. Hence, $\max \pi_i\left(b_{ij}(h), c_{iji}(h)\right) = \pi_i\left(30, c_{iji}(h)\right)$ and $\min \pi_i\left(b_{ij}(h), c_{iji}(h)\right) = \pi_i\left(0, c_{iji}(h)\right)$.

Given that $c_{iji}(h)$ is a probability distribution, then the payoff that person i beliefs that person j intends to give person i by contributing $b_{ij}(h) = g_j$ is an expected payoff of all the potential payoffs that person i could get for every action that person i makes weighted by the corresponding probability value in the probability distribution of $c_{iji}(h)$. In more intuitive terms, we can rewrite $\pi_i\left(g_j, c_{iji}(h)\right)$, $\pi_i\left(30, c_{iji}(h)\right)$ and $\pi_i\left(0, c_{iji}(h)\right)$ as follows:

$$\pi_i\left(g_j, c_{iji}(h)\right)$$
$$= p'' \times \pi_i\left(g_j, g_i = 0\right) + q'' \times \pi_i\left(g_j, g_i = 10\right) + r'' \times \pi_i\left(g_j, g_i = 20\right)$$
$$+ (1 - p'' - q'' - r'') \times \pi_i\left(g_j, g_i = 30\right)$$

$$\pi_i\left(30, c_{iji}(h)\right)$$
$$= p'' \times \pi_i(30, g_i = 0) + q'' \times \pi_i(30, g_i = 10) + r'' \times \pi_i(30, g_i = 20)$$
$$+ (1 - p'' - q'' - r'') \times \pi_i(30, g_i = 30)$$

$$\pi_i\left(0, c_{iji}(h)\right) = p'' \times \pi_i(0, g_i = 0) + q'' \times \pi_i(0, g_i = 10) + r'' \times \pi_i(0, g_i = 20)$$
$$+ (1 - p'' - q'' - r'') \times \pi_i(0, g_i = 30)$$

Substituting each of the relevant elements of the RHS in each of the previous three equations by the corresponding material payoff function described earlier, we get:

$$\pi_i\left(g_j, c_{iji}(h)\right)$$
$$= p'' \times \left(30 - 0 + \underline{m} \times \left(g_j + 0\right)\right) + q'' \times \left(30 - 10 + \underline{m} \times \left(g_j + 10\right)\right)$$
$$+ r'' \times \left(30 - 20 + \underline{m} \times \left(g_j + 20\right)\right)$$
$$+ (1 - p'' - q'' - r'') \times \left(30 - 30 + \underline{m} \times \left(g_j + 30\right)\right)$$

$$\pi_i\left(30, c_{iji}(h)\right)$$
$$= p'' \times \left(30 - 0 + \underline{m} \times (0 + 30)\right) + q'' \times \left(30 - 10 + \underline{m} \times (30 + 10)\right)$$
$$+ r'' \times \left(30 - 20 + \underline{m} \times (30 + 20)\right)$$
$$+ (1 - p'' - q'' - r'') \times \left(30 - 30 + \underline{m} \times (30 + 30)\right)$$

$$\pi_i\left(0, c_{iji}(h)\right) = p'' \times \left(30 + \underline{m} \times (0 + 0)\right) + q'' \times \left(20 + \underline{m} \times (0 + 10)\right)$$
$$+ r'' \times \left(30 - 20 + \underline{m} \times (0 + 20)\right)$$
$$+ (1 - p'' - q'' - r'') \times \left(30 - 30 + \underline{m} \times (0 + 30)\right)$$

Which simplify to:

$$\pi_i\left(g_j, c_{iji}(h)\right)$$
$$= p'' \times \left(30 + \underline{m} \times (g_j)\right) + q'' \times \left(20 + \underline{m} \times (g_j + 10)\right)$$
$$+ r'' \times \left(10 + \underline{m} \times (g_j + 20)\right) + (1 - p'' - q'' - r'') \times \left(\underline{m} \times (g_j + 30)\right)$$

$$\pi_i\left(30, c_{iji}(h)\right)$$
$$= p'' \times (30 + \underline{m} \times 30) + q'' \times (20 + \underline{m} \times 40) + r'' \times (10 + \underline{m} \times 50)$$
$$+ (1 - p'' - q'' - r'') \times (\underline{m} \times 60)$$

$$\pi_i\left(0, c_{iji}(h)\right) = p'' \times (30) + q'' \times (20 + \underline{m} \times 10) + r'' \times (10 + \underline{m} \times 20)$$
$$+ (1 - p'' - q'' - r'') \times (\underline{m} \times 30)$$

Using the last two equations, and taking p'', q'', r'' and $1 - p'' - q'' - r''$ as common factors, we can express $\pi_i\left(30, c_{iji}(h)\right) + \pi_i\left(0, c_{iji}(h)\right)$ as:

$$\pi_i\left(30, c_{iji}(h)\right) + \pi_i\left(0, c_{iji}(h)\right)$$
$$= p'' \times \left(30 + \underline{m} \times 30 + 30\right) + q'' \times \left(20 + \underline{m} \times 40 + 20 + \underline{m} \times 10\right)$$
$$+ r'' \times \left(10 + \underline{m} \times 50 + 10 + \underline{m} \times 20\right)$$
$$+ (1 - p'' - q'' - r'') \times \left(\underline{m} \times 60 + \underline{m} \times 30\right)$$

Which can be simplified to:

$$\pi_i\left(30, c_{iji}(h)\right) + \pi_i\left(0, c_{iji}(h)\right)$$
$$= p'' \times \left(60 + \underline{m} \times 30\right) + q'' \times \left(40 + \underline{m} \times 50\right) + r'' \times \left(20 + \underline{m} \times 70\right)$$
$$+ (1 - p'' - q'' - r'') \times \left(\underline{m} \times 90\right)$$

Hence, the second term of the perceived kindness function, $\frac{\pi_i\left(30, c_{iji}(h)\right) + \pi_i\left(0, c_{iji}(h)\right)}{2}$, can be written as:

$$\frac{\pi_i\left(30, c_{iji}(h)\right) + \pi_i\left(0, c_{iji}(h)\right)}{2}$$
$$= p'' \times \left(30 + \underline{m} \times 15\right) + q'' \times \left(20 + \underline{m} \times 25\right) + r'' \times \left(10 + \underline{m} \times 35\right)$$
$$+ (1 - p'' - q'' - r'') \times \underline{m} \times 45$$

Now, using the expressions we found for $\pi_i\left(g_j, c_{iji}(h)\right)$ and $\frac{\pi_i\left(30, c_{iji}(h)\right) + \pi_i\left(0, c_{iji}(h)\right)}{2}$, and taking p'', q'', r'' and $1 - p'' - q'' - r''$ as common factors, we can express the perceived kindness function as:

$$\lambda_{iji}\left(g_j, c_{iji}(h)\right)$$
$$= p'' \times \left(30 + \underline{m} \times g_j - 30 - \underline{m} \times 15\right)$$
$$+ q'' \times \left(20 + \underline{m} \times \left(g_j + 10\right) - 20 - \underline{m} \times 25\right)$$
$$+ r'' \times \left(10 + \underline{m} \times \left(g_j + 20\right) - 10 - \underline{m} \times 35\right)$$
$$+ (1 - p'' - q'' - r'') \times \left(\underline{m} \times \left(g_j + 30\right) - \underline{m} \times 45\right)$$

By taking $\underline{m}$ as a common factor and simplifying, we get:

$$\lambda_{iji}\left(g_j, c_{iji}(h)\right)$$
$$= p'' \times \left(\underline{m} \times (g_j - 15)\right) + q'' \times \left(\underline{m} \times (g_j + 10 - 25)\right)$$
$$+ r'' \times \left(\underline{m} \times (g_j + 20 - 35)\right)$$
$$+ (1 - p'' - q'' - r'') \times \left(\underline{m} \times (g_j + 30 - 45)\right)$$

Which can be further simplified to:

$$\lambda_{iji}\left(g_j, c_{iji}(h)\right)$$
$$= p'' \times \left(\underline{m} \times (g_j - 15)\right) + q'' \times \left(\underline{m} \times (g_j - 15)\right)$$
$$+ r'' \times \left(\underline{m} \times (g_j - 15)\right) + (1 - p'' - q'' - r'') \times \left(\underline{m} \times (g_j - 15)\right)$$

Now, taking $\underline{m} \times (g_j - 15)$ as a common factor, we can rewrite the previous expression as:

$$\lambda_{iji}\left(g_j, c_{iji}(h)\right) = \underline{m} \times (g_j - 15) \times (p'' + q'' + r'' + 1 - p'' - q'' - r'')$$

Which can be further simplified to:

$$\lambda_{iji}\left(g_j, c_{iji}(h)\right) = \underline{m} \times (g_j - 15)$$

Step 3: Substitute the two expressions found in the reciprocity utility function.

Given the expressions of the kindness and perceived kindness function of person i, we can rewrite his or her utility as:

$$U_i^{DK}(\pi_i, \pi_j) = \pi_i\left(g_i(h), g_j, \kappa_{ij}, \lambda_{iji}\right) = \pi_i\left(g_i, g_j\right) + Y_{i,j} \times \underline{m} \times (g_i - 15) \times \underline{m} \times (g_j - 15)$$

Which, substituting $\pi_i(g_i, g_j)$ by the payoff function given g_i and g_j, we get:

$$U_i^{DK}(\pi_i, \pi_j) = 30 - g_i + \underline{m} \times (g_i + g_j) + Y_{i,j} \times \underline{m}^2 \times (g_i - 15) \times (g_j - 15)$$

Step 4: Compute the first order derivative of the utility function.

Taking the first derivative of the utility function with respect to the contribution of person i, we get:

$$\frac{\partial U_i^{DK}(\pi_i, \pi_j)}{\partial g_i} = -1 + \underline{m} + Y_{i,j} \times \underline{m}^2 \times (g_j - 15)$$

Step 5: Compute the sign of first order derivative of the utility function for $g_j \in \{0{,}10\}$.

When $g_j \in \{0{,}10\}$, then $g_j - 15 = (\leq 10) - 15 = (< 0)$. As $Y_{i,j} > 0$, and $\underline{m} < 1$, it, hence, follows that:

$$\frac{\partial U_i^{DK}(\pi_i, \pi_j)}{\partial g_i} = -1 + (< 1) + (\geq 0) \times \underline{m}^2 \times (< 0) = (< 0) + (< 0) = (< 0)$$

Hence,

$$\frac{\partial U_i^{DK}(\pi_i, \pi_j)}{\partial g_i} < 0$$

Which demonstrates that the utility function is decreasing over the whole domain of g_i for $g_j \in \{0{,}10\}$.

Step 6: Compute the optimal contribution of person i *against* $g_j \in \{0{,}10\}$.

Given that, for $g_j \in \{0{,}10\}$, the derivative of the utility function is negative over the whole domain of g_i, person i will maximise their utility by contributing nothing. That is,

$$(\forall\, Y_{i,j}), c_i^* = g_i = 0 \forall g_j \in \{0{,}10\}$$

Step 7: Compute the sign of first order derivative of the utility function for $g_j \in \{20{,}30\}$ *in terms of* $Y_{i,j}$.

The marginal utility becomes negative iff:

$$-1 + \underline{m} + Y_{i,j} \times \underline{m}^2 \times (g_j - 15) < 0$$

Isolating $Y_{i,j}$ if the LHS, we get:

$$Y_{i,j} \times \underline{m}^2 \times (g_j - 15) < 1 - \underline{m}$$

Dividing both sides by $\underline{m}^2 \times (g_j - 15)$, we get:

$$Y_{i,j} < \frac{1 - \underline{m}}{\underline{m}^2 \times (g_j - 15)} \; iff \; \frac{\partial U_i^{DK}(\pi_i, \pi_j)}{\partial g_i} < 0$$

For $g_j \in \{20{,}30\}$, whenever $Y_{i,j}$ is lower than the threshold value found above, the marginal utility with respect to g_i will be negative. In contrast, whenever the marginal utility is positive, we get the following condition:

$$Y_{i,j} > \frac{1 - \underline{m}}{\underline{m}^2 \times (g_j - 15)} \; iff \; \frac{\partial U_i^{DK}(\pi_i, \pi_j)}{\partial g_i} < 0$$

And whenever the marginal utility is exactly 0, it then follows that:

$$Y_{i,j} = \frac{1 - \underline{m}}{\underline{m}^2 \times (g_j - 15)} \; iff \; \frac{\partial U_i^{DK}(\pi_i, \pi_j)}{\partial g_i} = 0$$

Step 8: Compute the optimal contribution of person i *against* $g_j \in \{20{,}30\}$ *for all possible values of* $Y_{i,j}$.

Given the inequalities found in the previous step, the best responses against $g_j \in \{20,30\}$ can be summarised as:

$$c_i^* = \begin{cases} g_i = 0\ \forall g_j \in \{20,30\}\ iff\ Y_{i,j} < \dfrac{1-\underline{m}}{\underline{m}^2 \times (g_j - 15)} \\ g_i \in A_i\ \forall g_j \in \{20,30\}\ iff\ Y_{i,j} = \dfrac{1-\underline{m}}{\underline{m}^2 \times (g_j - 15)} \\ g_i = 30\ \forall g_j \in \{20,30\}\ iff\ Y_{i,j} > \dfrac{1-\underline{m}}{\underline{m}^2 \times (g_j - 15)} \end{cases}$$

Where the previous results hold given the linearity of the utility function $U_i^{DK}(\pi_i, \pi_j)$. That is, whenever the derivative is decreasing in the whole domain of g_i, as it is the case of the first of the two equations, then the best answer is to free ride; and whenever the derivative is increasing in the whole domain of g_i, as is the case of the second of the two equations, the best answer is to fully contribute. Whenever the derivative is equal to zero, any contribution gives the same utility and hence all are optimal choices. The sign of the derivative is determined by the reciprocity parameter $Y_{i,j}$.

Step 9: show that only full contribution (i.e., $g_i = 30$) is an efficient strategy in the CIG.

Unlike in the SDG, now only full contribution is an efficient strategy in a common interest game. This is the case as, for each and every of the contributions of the first mover player j – that is, for each of the possible histories of play before player i gets to play –, full contribution by player i gives no lower material payoff to any player and a higher material payoff to all players. As Player i's contribution decision is the only subsequent play for each and every contribution of player j, then by Dufwenberg and Kirchsteiger's (2004, pp. 276) definition of the set of efficient strategies, it follows that full contribution is the only strategy within the set of efficient strategies of player i, $E_i = \{g_i = 30\}$.

To see why $g_i = 30$ gives no lower material payoff to any of the players, notice that, in a common interest game, $\overline{m} \in (1, \infty)$. Hence, start by assuming that $g_i = 30$ implies

$$\pi_i(30, g_j) > \pi_i\left(\underline{g_i}, g_j\right)$$

Substituting the material payoff function by its functional form yields:

$$\overline{m} \times (30 + g_j) > 30 - \underline{g_i} + \overline{m} \times \left(\underline{g_i} + g_j\right)$$

Where $\underline{g_i} < 30$ is an arbitrarily small contribution of player i. Bringing $\overline{m}$ to the LHS, and taking $\overline{m}$ as a common factor, we get:

$$\overline{m} \times \left(30 + g_j - \underline{g_i} - g_j\right) > 30 - \underline{g_i}$$

Simplifying the parenthesis in the LHS, we get:

$$\overline{m} \times \left(30 - \underline{g_i}\right) > 30 - \underline{g_i}$$

Dividing both hand sides by $\left(30 - \underline{g_i}\right)$, we get:

$$\overline{m} > 1$$

Which is exactly the condition that will always hold in common interest games, thereby discharging the initial assumption. Hence, it follows that $g_i = 30$ gives the highest material payoff to player i.

Now, consider the payoff function of player j:

$$\pi_j(g_i, g_j) = 30 - g_j + \overline{m} \times (g_i + g_j)$$

The derivative of the function with respect to g_i is given by:

$$\frac{\partial \pi_j(g_i, g_j)}{\partial g_j} = \overline{m}$$

As $\overline{m} \in (1, \infty)$ in common interest games, it follows that $\frac{\partial \pi_j(g_i, g_j)}{\partial g_j} = \overline{m} > 0$. As the payoff function is linear in the contribution of player i and it is also increasing in it, it follows that $g_i = 30$ is the contribution of player i that will maximise the payoff of player j.
Hence, it follows that there doesn't exist another g_i that gives a higher payoff to any of the players, thereby proving why $g_i = 30$ is the only efficient strategy in common interest games.

Step 10: Outline the implications of a reduced set of efficient strategies in the kindness function (κ_{ij}) and the perceived kindness function (λ_{iji}) of subject i in the CIG.

This has important implications when computing the equitable payoff in both the kindness and perceived kindness functions, as the minimum payoff that can be given to any player is evaluated within the strategies that are efficient. Hence,

$$\begin{aligned} \min \pi_j\big(g_i, b_{ij}(h) = g_j\big)|g_i \in E_i &= \max \pi_j\big(g_i, b_{ij}(h) = g_j\big)|g_i \in A_i \\ &= \pi_j\big(g_i = 30, b_{ij}(h) = g_j\big) \end{aligned}$$

and

$$\max \pi_i \left(b_{ij}(h) = g_j, c_{iji}(h)\right) |g_j \in A_j = \min \pi_i \left(b_{ij}(h) = g_j, c_{iji}(h)\right) |g_j \in E_j = \pi_i \left(b_{ij}(h) = 30, c_{iji}(h)\right).$$

The implication for the kindness and perceived kindness functions is that they can be defined as:

$$\kappa_{ij}\left(g_i, b_{ij}(h)\right) = \pi_j\big(g_i, g_j\big) - \frac{2 \times \pi_j\big(30, g_j\big)}{2}$$

$$\lambda_{iji}\left(b_{ij}(h), c_{iji}(h)\right) = \pi_i\left(g_j, c_{iji}(h)\right) - \frac{2 \times \pi_i\left(30, c_{iji}(h)\right)}{2}$$

Which can be simplified to:

$$\kappa_{ij}\left(g_i, b_{ij}(h)\right) = \pi_j\big(g_i, g_j\big) - \pi_j\big(30, g_j\big)$$

$$\lambda_{iji}\left(b_{ij}(h), c_{iji}(h)\right) = \pi_i\left(g_j, c_{iji}(h)\right) - \pi_i\left(30, c_{iji}(h)\right)$$

Step 11: find the kindness function (κ_{ij}) of subject i in the CIG.

At generic contribution levels g_j and g_i, then $b_{ij}(h) = g_j$. Hence, we can write the kindness function as:

$$\kappa_{ij}\left(g_{ij}(h), b_{ij}(h)\right) = \pi_j(g_i, g_j) - \pi_j(30, g_j)$$

Substituting $\pi_j(g_i, g_j)$ by the payoff function outlined above, we get:

$$\kappa_{ij}\left(g_{ij}(h), b_{ij}(h)\right) = 30 - g_j + \overline{m} \times (g_i + g_j) - 30 + g_j - \overline{m} \times (30 + g_j)$$

Simplifying, we get:

$$\kappa_{ij}\left(g_{ij}(h), b_{ij}(h)\right) = \overline{m} \times (g_i + g_j) - \overline{m} \times (30 + g_j)$$

Using $\overline{m}$ as a common factor, we get:

$$\kappa_{ij}\left(g_{ij}(h), b_{ij}(h)\right) = \overline{m} \times (g_i + g_j - 30 - g_j)$$

And, finally, simplifying we get:

$$\kappa_{ij}\left(g_{ij}(h), b_{ij}(h)\right) = \overline{m} \times (g_i - 30)$$

Step 12: find the perceived kindness function (λ_{iji}) of subject i in the CIG.

We can define the perceived kindness function of player i as:

$$\lambda_{iji}\left(b_{ij}(h), c_{iji}(h)\right) = \pi_i\left(g_j, c_{iji}(h)\right) - \pi_i\left(30, c_{iji}(h)\right)$$

Given that $c_{iji}(h)$ is the probability distribution described earlier, we can rewrite $\pi_i\left(g_j, c_{iji}(h)\right)$ and $\pi_i\left(30, c_{iji}(h)\right)$ as follows:

$$\begin{aligned}\pi_i\left(g_j, c_{iji}(h)\right) \\ &= p'' \times \pi_i(g_j, g_i = 0) + q'' \times \pi_i(g_j, g_i = 10) + r'' \times \pi_i(g_j, g_i = 20) \\ &+ (1 - p'' - q'' - r'') \times \pi_i(g_j, g_i = 30)\end{aligned}$$

$$\begin{aligned}\pi_i\left(30, c_{iji}(h)\right) \\ &= p'' \times \pi_i(30, g_i = 0) + q'' \times \pi_i(30, g_i = 10) + r'' \times \pi_i(30, g_i = 20) \\ &+ (1 - p'' - q'' - r'') \times \pi_i(30, g_i = 30)\end{aligned}$$

Substituting each of the elements of the RHS in each of the previous three equations by the corresponding payoff function described earlier, we get:

$$\begin{aligned}\pi_i\left(g_j, c_{iji}(h)\right) \\ &= p'' \times \left(30 - 0 + \overline{m} \times (g_j + 0)\right) + q'' \times \left(30 - 10 + \overline{m} \times (g_j + 10)\right) \\ &+ r'' \times \left(30 - 20 + \overline{m} \times (g_j + 20)\right) \\ &+ (1 - p'' - q'' - r'') \times \left(30 - 30 + \overline{m} \times (g_j + 30)\right)\end{aligned}$$

$$\begin{aligned}\pi_i\left(30, c_{iji}(h)\right) \\ &= p'' \times \left(30 - 0 + \overline{m} \times (0 + 30)\right) + q'' \times \left(30 - 10 + \overline{m} \times (30 + 10)\right) \\ &+ r'' \times \left(30 - 20 + \overline{m} \times (30 + 20)\right) \\ &+ (1 - p'' - q'' - r'') \times \left(30 - 30 + \overline{m} \times (30 + 30)\right)\end{aligned}$$

Which simplify to:

$$\pi_i\left(g_j, c_{iji}(h)\right)$$
$$= p'' \times \left(30 + \overline{m} \times \left(g_j\right)\right) + q'' \times \left(20 + \overline{m} \times \left(g_j + 10\right)\right)$$
$$+ r'' \times \left(10 + \overline{m} \times \left(g_j + 20\right)\right) + (1 - p'' - q'' - r'') \times \left(\overline{m} \times \left(g_j + 30\right)\right)$$

$$\pi_i\left(30, c_{iji}(h)\right)$$
$$= p'' \times (30 + \overline{m} \times 30) + q'' \times (20 + \overline{m} \times 40) + r'' \times (10 + M\overline{m} \times 50)$$
$$+ (1 - p'' - q'' - r'') \times (\overline{m} \times 60)$$

Now, using the expressions we found for $\pi_i\left(g_j, c_{iji}(h)\right)$ and $\pi_i\left(30, c_{iji}(h)\right)$, and taking p'', q'', r'' and $1 - p'' - q'' - r''$ as common factors, we can express the perceived kindness function as:

$$\lambda_{iji}\left(b_{ij}(h), c_{iji}(h)\right)$$
$$= p'' \times \left(30 + \overline{m} \times g_j - 30 - \overline{m} \times 30\right)$$
$$+ q'' \times \left(20 + \overline{m} \times \left(g_j + 10\right) - 20 - \overline{m} \times 40\right)$$
$$+ r'' \times \left(10 + \overline{m} \times \left(g_j + 20\right) - 10 - \overline{m} \times 50\right)$$
$$+ (1 - p'' - q'' - r'') \times \left(\overline{m} \times \left(g_j + 30\right) - \overline{m} \times 60\right)$$

By taking $\overline{m}$ as a common factor and simplifying, we get:

$$\lambda_{iji}\left(b_{ij}(h), c_{iji}(h)\right)$$
$$= p'' \times \left(\overline{m} \times \left(g_j - 30\right)\right) + q'' \times \left(\overline{m} \times \left(g_j + 10 - 40\right)\right)$$
$$+ r'' \times \left(\overline{m} \times \left(g_j + 20 - 50\right)\right)$$
$$+ (1 - p'' - q'' - r'') \times \left(\overline{m} \times \left(g_j + 30 - 60\right)\right)$$

Which can be further simplified to:

$$\lambda_{iji}\left(b_{ij}(h), c_{iji}(h)\right)$$
$$= p'' \times \left(\overline{m} \times \left(g_j - 30\right)\right) + q'' \times \left(\overline{m} \times \left(g_j - 30\right)\right)$$
$$+ r'' \times \left(\overline{m} \times \left(g_j - 30\right)\right) + (1 - p'' - q'' - r'') \times \left(\overline{m} \times \left(g_j - 30\right)\right)$$

Now, taking $\overline{m} \times \left(g_j - 30\right)$ as a common factor, we can rewrite the previous expression as:

$$\lambda_{iji}\left(b_{ij}(h) = g_j, c_{iji}(h)\right) = \overline{m} \times \left(g_j - 30\right) \times (p'' + q'' + r'' + 1 - p'' - q'' - r'')$$

Which can be further simplified to:

$$\lambda_{iji}\left(b_{ij}(h), c_{iji}(h)\right) = \overline{m} \times \left(g_j - 30\right)$$

Step 13: Substitute the two expressions found in the reciprocity utility function.

Given the expressions of the kindness and perceived kindness function of person i, we can rewrite his or her utility as:

$$U_i^{DK}(\pi_i, \pi_j) = \pi_i\left(g_i, g_j, b_{ij}(h), c_{iji}(h)\right)$$
$$= \pi_i\left(g_i, b_{ij}(h)\right) + Y_{i,j} \times \overline{m} \times (g_i - 30) \times \overline{m} \times \left(g_j - 30\right)$$

Which, substituting $\pi_i\left(g_i, b_{ij}(h)\right)$ by the material payoff function given g_i and g_j, for a generic first-order belief of $g_j$we get:

$$U_i^{DK}(\pi_i, \pi_j) = 30 - g_i + \overline{m} \times \left(g_i + g_j\right) + Y_{i,j} \times \overline{m}^2 \times (g_i - 30) \times \left(g_j - 30\right)$$

Step 13: find the first order derivative of the utility function with respect to g_i.

Taking the first derivative of the utility function with respect to the contribution of person i, we get:

$$\frac{\partial U_i^{DK}(\pi_i, \pi_j)}{\partial g_i} = -1 + \overline{m} + Y_{i,j} \times \overline{m}^2 \times (g_j - 30)$$

Step 14: find the optimal contribution for person i *against* $g_j = 30$.

Note that, whenever $g_j = 30$, then $g_j - 30 = 0$. Hence, the reciprocal term collapses to 0 regardless of the value of $Y_{i,j}$. Hence, when $g_j = 30$ the marginal utility of own contribution is given by:

$$\left.\frac{\partial U_i^{DK}(\pi_i, \pi_j)}{\partial g_i}\right|_{g_j=30} = -1 + \overline{m}$$

As $\overline{m} > 1$ it follows that the marginal utility of own contribution when $g_j = 30$ will always be positive:

$$\left.\frac{\partial U_i^{DK}(\pi_i, \pi_j)}{\partial g_i}\right|_{g_j=30} = -1 + (> 1) = (> 0)$$

This implies that the best response against $g_j = 30$, given the linearity of the utility function with respect to own contribution, will be

$$(\forall\, Y_{i,j}), c_i^* = g_i = 30 \; if \; g_j = 30$$

Step 15: find the optimal contribution for person i *against* $g_j \in \{0,10,20\}$.

Turning to the remaining cases, that is $g_j \in \{0,10,20\}$, we need to find for which values of $Y_{i,j}$ the marginal utility becomes negative. Recalling the marginal utility of g_i, we can capture that case with the following inequality:

$$-1 + \overline{m} + Y_{i,j} \times \overline{m}^2 \times (g_j - 30) < 0$$

Isolating $Y_{i,j}$ in the RHS, we get:

$$Y_{i,j} \times \overline{m}^2 \times (30 - g_j) > \overline{m} - 1$$

Dividing both sides by $\overline{m}^2 \times (30 - g_j)$, we get:

$$Y_{i,j} > \frac{\overline{m} - 1}{\overline{m}^2 \times (30 - g_j)}$$

For $g_j \in \{0,10,20\}$, then, we can capture person i's best responses as:

$$c_i^* = \begin{cases} g_i = 0 \ \forall g_j \in \{0,10,20\} \ iff \ Y_{i,j} > \dfrac{\overline{m} - 1}{\overline{m}^2 \times (30 - g_j)} \\ g_i \in A_i \ \forall g_j \in \{0,10,20\} \ iff \ Y_{i,j} = \dfrac{\overline{m} - 1}{\overline{m}^2 \times (30 - g_j)} \\ g_i = 30 \ \forall g_j \in \{0,10,20\} \ iff \ Y_{i,j} > \dfrac{\overline{m} - 1}{\overline{m}^2 \times (30 - g_j)} \end{cases}$$

QED.

A.2.4.3. Other results involving reciprocity preferences

We use the results from proposition 3 to provide, in corollary 3.1, the precise contribution attitudes in the SDG and CIG that we use in chapter 4. Additionally, we provide another main result besides proposition 3. Namely, that for some joint values of $\underline{m}$ and $\overline{m}$ person i cannot be a conditional cooperator in the SDG without being a conditional cooperator in the CIG. Hence, for such values of $\underline{m}$ and $\overline{m}$ preferences for reciprocity cannot predict conditional cooperation in the SDG and unconditional cooperation in the CIG. We summarise this statement in corollary3.2. Additionally, corollary 3.3 shows that, for the values of $\underline{m}$ and $\overline{m}$ used in the experiments of chapter 4, the result from corollary 3.2 holds true in our data. That is, preferences for reciprocity cannot rationalise conditional cooperation in the SDG and unconditional cooperation in the CIG.

Corollary 2.1. *If subject i maximizes the utility function* $U_i^{DK}\left(\pi_i(g_i, g_j), \pi_j(g_i, g_j)\right)$, *where i contributes* g_i, *the other player contributes* g_j, *and the other player moves first and subject i second, and where we denote* c_i^* *as subject i's optimal contribution schedule, then subject i will*

(i), in the Social Dilemma,

- *(a) do* $c_i^* = g_i = 0$ *against* $g_j \in \{0,10\}$ *regardless of* $Y_{i,j}$
- *(b) do* $c_i^* = g_i = 0$ *against* $g_j = 20$ *iff* $Y_{i,j} < \frac{0.4}{1.8}$
- *(c) do* $c_i^* = g_i = 30$ *against* $g_j = 20$ *iff* $Y_{i,j} > \frac{0.4}{1.8}$
- *(d) do* $c_i^* = g_i = 0$ *against* $g_j = 30$ *iff* $Y_{i,j} < \frac{0.4}{5.4}$
- *(e) do* $c_i^* = g_i = 30$ *against* $g_j = 30$ *iff* $Y_{i,j} > \frac{0.4}{5.4}$

(ii), in the Common Interest Game,

- *(f) do* $c_i^* = g_i = 30$ *against* $g_j = 30$ *regardless of* $Y_{i,j}$
- *(g) do* $c_i^* = g_i = 0$ *against* $g_j = 0$*) iff* $Y_{i,j} > \frac{0.2}{1.2^2 \times (30)}$
- *(h) do* $c_i^* = g_i = 30$ *against* $g_j = 0$*) iff* $Y_{i,j} < \frac{0.2}{1.2^2 \times (30)}$
- *(i) do* $c_i^* = g_i = 0$ *against* $g_j = 10$*) iff* $Y_{i,j} > \frac{0.2}{1.2^2 \times (20)}$
- *(j) do* $c_i^* = g_i = 30$ *against* $g_j = 10$*) iff* $Y_{i,j} < \frac{0.2}{1.2^2 \times (20)}$
- *(k) do* $c_i^* = g_i = 0$ *against* $g_j = 20$*) iff* $Y_{i,j} > \frac{0.2}{1.2^2 \times (10)}$
- *(l) do* $c_i^* = g_i = 30$ *against* $g_j = 20$*) iff* $Y_{i,j} < \frac{0.2}{1.2^2 \times (10)}$

Proof.

Given the contribution attitudes found in proposition 3, (a) and (f) follow without further demonstration. Substituting $\underline{m} = 0.6$ in the cooperation attitudes found in proposition 3, we get the following expressions for the SDG:

$c_i^* = g_i = 0$ against $g_j \in \{20,30\}$ iff $Y_{i,j} < \frac{1-0.6}{0.6^2 \times (g_j - 15)}$

$c_i^* = g_i = 30$ against $g_j \in \{20,30\}$ iff $Y_{i,j} > \frac{1-0.6}{0.36 \times (g_j - 15)}$

Substituting g_j explicitly in the inequalities, we get:

$c_i^* = g_i = 0$ against $g_j = 20$ iff $Y_{i,j} < \frac{1-0.6}{0.36\times(5)}$

$c_i^* = g_i = 30$ against $g_j = 20$ iff $Y_{i,j} > \frac{1-0.6}{0.36\times(5)}$

$c_i^* = g_i = 0$ against $g_j = 30$ iff $Y_{i,j} < \frac{1-0.6}{0.36\times(15)}$

$c_i^* = g_i = 30$ against $g_j = 30$ iff $Y_{i,j} > \frac{1-0.6}{0.36\times(15)}$

And, simplifying, we get:

$c_i^* = g_i = 0$ against $g_j = 20$ iff $Y_{i,j} < \frac{0.4}{1.8}$

$c_i^* = g_i = 30$ against $g_j = 20$ iff $Y_{i,j} > \frac{0.4}{1.8}$

$c_i^* = g_i = 0$ against $g_j = 30$ iff $Y_{i,j} < \frac{0,4}{5.4}$

$c_i^* = g_i = 30$ against $g_j = 30$ iff $Y_{i,j} > \frac{0.4}{5.4}$

Which proves (b), (c), (d), and (e). Additionally, substituting $\overline{m} = 1.2$ in the cooperation attitudes found in proposition 3, we get the following expressions for the CIG:

$c_i^* = g_i = 0$ against $g_j \in \{0,10,20\})$ iff $Y_{i,j} > \frac{1.2-1}{1.2^2\times(30-g_j)}$

$c_i^* = g_i = 30$ against $g_j \in \{0,10,20\})$ iff $Y_{i,j} < \frac{1.2-1}{1.2^2\times(30-g_j)}$

Substituting g_j explicitly in the inequalities, we get:

$c_i^* = g_i = 0$ against $g_j = 0$ iff $Y_{i,j} > \frac{1.2-1}{1.2^2\times(30)}$

$c_i^* = g_i = 30$ against $g_j = 0$ iff $Y_{i,j} < \frac{1.2-1}{1.2^2\times(30)}$

$c_i^* = g_i = 0$ against $g_j = 10$ iff $Y_{i,j} > \frac{1.2-1}{1.2^2\times(20)}$

$c_i^* = g_i = 30$ against $g_j = 10$ iff $Y_{i,j} < \frac{1.2-1}{1.2^2\times(20)}$

$c_i^* = g_i = 0$ against $g_j = 20$ iff $Y_{i,j} > \frac{1.2-1}{1.2^2\times(10)}$

$c_i^* = g_i = 30$ against $g_j = 20$ iff $Y_{i,j} < \frac{1.2-1}{1.2^2 \times (10)}$

And, simplifying, we get:

$c_i^* = g_i = 0$ against $g_j = 0$ iff $Y_{i,j} > \frac{0.2}{1.2^2 \times (30)}$

$c_i^* = g_i = 30$ against $g_j = 0$ iff $Y_{i,j} < \frac{0.2}{1.2^2 \times (30)}$

$c_i^* = g_i = 0$ against $g_j = 10$ iff $Y_{i,j} > \frac{0.2}{1.2^2 \times (20)}$

$c_i^* = g_i = 30$ against $g_j = 10$ iff $Y_{i,j} < \frac{0.2}{1.2^2 \times (20)}$

$c_i^* = g_i = 0$ against $g_j = 20$ iff $Y_{i,j} > \frac{0.2}{1.2^2 \times (10)}$

$c_i^* = g_i = 30$ against $g_j = 20$ iff $Y_{i,j} < \frac{0.2}{1.2^2 \times (10)}$

Which proves (g), (h), (i), (j), (k), and (l).

QED.

Corollary 2.2. *If subject* i *maximizes the utility function* $U_i^{DK}\left(\pi_i(g_i, g_j), \pi_j(g_i, g_j)\right)$, *where* i *contributes* g_i *and the other player contributes* g_j, *then if*

(i) *person* i *plays the weakest form of conditional cooperation possible in the SDG, and*

(ii) *it comes to pass that* $\frac{1-\underline{m}}{\underline{m}^2 \times (15)} > \frac{\overline{m}-1}{30 \times \overline{m}^2}$,

then subject i *must play at least the weakest form of conditional cooperation in the CIG.*

Proof.

Given proposition 3, the weakest conditional cooperation pattern predicted by reciprocity in the SDG entails subject i to fully contribute against full contribution and free ride otherwise. More formally, it entails subject i to play $c_i^* = g_i = 0$ against $g_j = \{0,10,20\}$ and $c_i^* = g_i = 30$ against $g_j = 30$ in the SDG. Also, the weakest form of conditional cooperation in the CIG entails free riding against free riding and full contribution otherwise. More formally, it entails subject i to play $c_i^* = g_i = 0$ against $g_j = 0$ and $c_i^* = g_i = 30$ against $g_j \in \{10,20,30\}$ in the CIG.

Given proposition 3, the referred pattern of cooperation attitude in the SDG holds iff:

$$Y_{i,j} > \frac{1-\underline{m}}{\underline{m}^2 \times (15)}$$

Then, given that $Y_{i,j} > \frac{1-\underline{m}}{\underline{m}^2\times(15)}$ and that condition (ii) entails $\frac{1-\underline{m}}{\underline{m}^2\times(15)} > \frac{\overline{m}-1}{30\times\overline{m}^2}$, it naturally follows that:

$$Y_{i,j} > \frac{1-\underline{m}}{\underline{m}^2 \times (15)} > \frac{\overline{m}-1}{30 \times \overline{m}^2} \rightarrow Y_{i,j} > \frac{\overline{m}-1}{30 \times \overline{m}^2}$$

Recall that, given proposition 3, it follows that playing $c_i^* = g_i = 0$ against $g_j = 0$ and $c_i^* = g_i = 30$ against $g_j \in \{10,20,30\}$ in the CIG reveals the following inequality regarding $Y_{i,j}$:

$$Y_{ij} > \frac{\overline{m}-1}{\overline{m}^2 \times (30 - g_j)}$$

Hence, it follows that for a subject maximizing U_i^{DK}, playing the weakest form of conditional cooperation in the SDG implies at least some conditional cooperation in the CIG.

QED.

Corollary 2.3. *Given* $\underline{m} = 0.6$ *and* $\overline{m} = 1.2$*, then the weakest form of conditional cooperation in the SDG implies at least a form of conditional cooperation in the CIG.*

Proof.

Recall from corollary 2.2 that, given the weakest form of conditional cooperation, if $\frac{1-\underline{m}}{\underline{m}^2\times(15)} > \frac{\overline{m}-1}{30\times\overline{m}^2}$ then reciprocity would predict conditional cooperation in the CIG. Substituting $\underline{m} = 0.6$ and $\overline{m} = 1.2$ in that condition, we get:

$$\frac{1-0.6}{0.36\times(15)} > \frac{1.2-1}{30\times 1.2^2}$$

Which can be rearranged and simplified so as to read:

$$0.8\times 1.2^2 > 0.072$$

As $1.2^2 > 1$, then it follows that $0.8\times(>1) = (>0.8)$. And, hence, as $(>0.8) > 0.072$, given $\underline{m} = 0.6$ and $\overline{m} = 1.2$ the weakest form of conditional cooperation in the SDG implies a form of conditional cooperation in the CIG.

QED.

A.2.5. Spiteful preferences

A.2.5.1. Proof of proposition 4

Let's assume a subject's utility function, given g_i and g_j, is:

$$U_i^S(g_i,g_j) = \begin{cases} 30 - g_i + m\times(g_i+g_j) - \beta_i\times(g_j-g_i)\ if\ g_i \le g_j \\ 30 - g_i + m\times(g_i+g_j)\ if\ g_i \ge g_j \end{cases}$$

Where $\beta_i \le 0$. That is, a person with these preferences feels either pleasure or is indifferent at advantageous inequality ($\frac{\partial U_i(g_i,g_j)}{\partial(g_j-g_i)} = -\beta_i \ge 0$). These preferences represent someone who (i) derives pleasure from inequality provided that he is the one being better off in the distribution outcome. Otherwise, he does not feel any disadvantageous inequality. This is just the spiteful utility function $U_i^S(\pi_i,\pi_j)$ presented in chapter 4 once we substitute the material payoff function of the public goods game we are analysing.

Proposition 4. *If subject i maximizes the utility function $U_i^S\left(\pi_i(g_i, g_j), \pi_j(g_i, g_j)\right)$, where i contributes g_i and the other player contributes g_j, then subject i's contribution attitudes, denoted as c_i^*, will be*

(i), in the Social Dilemma,

$$(\forall \beta_i), c_i^* = g_i = 0 \; \forall g_j \in A_j$$

(ii), in the Common Interest Game,

$$c_i^* = \begin{cases} g_i = 30 \; if \; g_j = 0 & \forall \beta_i \\ g_i = 0 \; \forall g_j \in \{10,20,30\} & iff \; \beta_i < \dfrac{30 \times (1 - \overline{m})}{g_j} \\ g_i = 30 \; \forall g_j \in \{10,20,30\} & iff \; \beta_i > \dfrac{30 \times (1 - \overline{m})}{g_j} \end{cases}$$

Proof.

The marginal derivative with respect to own contributions is:

$$\frac{\partial U_i^S\left(\pi_i(g_i, g_j), \pi_j(g_i, g_j)\right)}{\partial g_i} = \begin{cases} -1 + m + \beta_i \; if \; g_i \leq g_j \\ -1 + m \; if \; g_i \geq g_j \end{cases}$$

For $\underline{m} < 1$, the second step of the marginal utility of own contributions is always negative. To see this, note $-1 + (< 1) = (< 0)$. The first step is negative when $\beta_i < 1 - \underline{m}$. For $\underline{m}$, it follows that $\beta_i < 1 - (< 1)$, as in the spiteful preferences model $\beta_i < 0$ and $1 - (< 1) = (> 0)$. Hence, the first derivative will be negative for all the values of $\underline{m} \in \left(\frac{1}{n}, 1\right)$. Given that the utility is linear in g_i and that the first derivative is negative alongside the whole domain of g_i for all values of $\underline{m}$, it follows that i's optimal cooperation attitudes in the SDG are given by:

$$(\forall \beta_i), c_i^* = g_i = 0 \; \forall \; g_j \in A_j$$

Which proves (i).

With regards to the CIG, the marginal derivative with respect to g_i is:

$$\frac{\partial U_i^S\left(\pi_i(g_i,g_j),\pi_j(g_i,g_j)\right)}{\partial g_i}=\begin{cases}-1+\overline{m}+\beta_i & if\ g_i<g_j\\ -1+\overline{m} & if\ g_i\geq g_j\end{cases}$$

For $\overline{m}\in(1,\infty)$, the second step of the marginal utility of own contributions is always positive. To see this, note that $\overline{m}>1$. Hence, $-1+(>1)=(>0)$. When $g_j=0$, then $g_i\geq 0$. Hence, against $g_j=0$ the best response is to fully contribute regardless of the value of β_i, as only the second step of the marginal derivative comes into play. This proves the first step of c_i^* in (ii). Notice that the first step of the marginal derivative is negative when $\beta_i<1-\overline{m}$ and positive when $\beta_i>1-\overline{m}$.

This implies that, whenever $\beta_i>1-\overline{m}$, both steps of the marginal utility will be positive and, hence, full contribution against all contributions of the other player will be the best response, as the marginal derivative will be positive alongside the whole domain of g_i. Hence, it follows that

$$c_i^*=g_i=30\ \forall g_j\in\{10,20,30\}\qquad\qquad iff\ \beta_i>\frac{30\times(1-\overline{m})}{g_j}$$

Thereby proving the last step in c_i^* of (ii).

Additionally, notice that, whenever $\beta_i<1-\overline{m}$, the first step of the marginal utility is negative. This implies that increasing contributions on the range $g_i<g_j$ decreases utility, thereby suggesting free riding as one potential optimal solution. The second step makes the marginal utility increasing in the range $g_i\geq g_j$, thereby suggesting full contribution as another potential optimal solution. Taken both results together, this indicates that we have two potential optimal best responses: free riding and full contribution.

Hence, person i's utility will be maximised by full contribution when $U_i^S(g_i=30,g_j)>U_i^S(g_i=0,g_j)$, which implies:

$$0 + \overline{m} \times (g_j + 30) > 30 + \overline{m} \times (g_j) - \beta_i \times (g_j)$$

Isolating β_i in the LHS and simplifying, we get:

$$\beta_i \times g_j > 30 + \overline{m} \times g_j - \overline{m} \times (g_j + 30)$$

Expanding the parenthesis of the RHS, we get:

$$\beta_i \times g_j > 30 + \overline{m} \times g_j - \overline{m} \times g_j - \overline{m} \times 30$$

Which, after simplifying, becomes:

$$\beta_i \times g_j > 30 - \overline{m} \times 30$$

And, taking 30 as a common factor in the RHS, we can rewrite the previous expression as:

$$\beta_i \times g_j > 30 \times (1 - \overline{m})$$

And, hence,

$$\beta_i > \frac{30 \times (1 - \overline{m})}{g_j}$$

Whenever $g_j > 0$ and $\beta_i < 1 - \overline{m}$, $U_i^S(g_i = 30, g_j) > U_i^S(g_i = 0, g_j)$ will hold true whenever $\beta_i > \frac{30\times(1-MPCR)}{g_j}$, and $U_i^S(g_i = 30, g_j) < U_i^S(g_i = 0, g_j)$ whenever $\beta_i < \frac{30\times(1-MPCR)}{g_j}$. Therefore, the optimal contributions given the values of β_i are:

$$c_i^* = g_i = 30\ \forall g_j \in \{10,20,30\} \qquad iff\ \beta_i > \frac{30 \times (1 - \overline{m})}{g_j}$$

$$c_i^* = g_i = 0\ \forall g_j \in \{10,20,30\} \qquad iff\ \beta_i < \frac{30 \times (1 - \overline{m})}{g_j}$$

Which finishes proving (ii).

QED.

A.2.5.2. Other results involving spiteful preferences

Below we provide a corollary that presents the specific threshold values of β_i determining optimal contributions for each g_j.

Corollary 4.1. *If subject i maximizes the utility function* $U_i^S\left(\pi_i(g_i, g_j), \pi_j(g_i, g_j)\right)$*, and given* $\underline{m} = 0.6$ *and* $\overline{m} = 1.2$*, the subject i's choices will*

(i), in the Social Dilemma, be

$$(\forall \beta_i), c_i^* = g_i = 0 \ \forall g_j \in A_j$$

(ii), in the Common Interest Game, be

(a) $(\forall \beta_i), g_i = 30 \ if \ g_j = 0$
(b) $g_i = 0$ *against* $g_j = 10$ *if* $\beta_i < -0.6$
(c) $g_i = 30$ *against* $g_j = 10$ *if* $\beta_i > -0.6$
(d) $g_i = 0$ *against* $g_j = 20$ *if* $\beta_i < -0.3$
(e) $g_i = 30$ *against* $g_j = 20$ *if* $\beta_i > -0.3$
(f) $g_i = 0$ *against* $g_j = 30$ *if* $\beta_i < -0.2$
(g) $g_i = 30$ *against* $g_j = 30$ *if* $\beta_i > -0.2$

Proof.

Part (i) trivially follows from proposition 4, and therefore needs no proof.

Regarding part (ii), recall the last two conditions found in proposition 4:

$$c_i^* = g_i = 30 \; \forall g_j \in \{10{,}20{,}30\} \qquad iff \; \beta_i > \frac{30 \times (1-\overline{m})}{g_j}$$

$$c_i^* = g_i = 0 \; \forall g_j \in \{10{,}20{,}30\} \qquad iff \; \beta_i < \frac{30 \times (1-\overline{m})}{g_j}$$

Substituting $\overline{m} = 1.2$ and simplifying, we get:

$$c_i^* = g_i = 30 \; \forall g_j \in \{10{,}20{,}30\} \qquad iff \; \beta_i > \frac{-6}{g_j}$$

$$c_i^* = g_i = 0 \; \forall g_j \in \{10{,}20{,}30\} \qquad iff \; \beta_i < \frac{-6}{g_j}$$

Substituting for all values of $g_i \in \{10{,}20{,}30\}$, we get the following conditions:

$c_i^* = g_i = 0$ against $g_j = 10$ *iff* $\beta_i < -0.6$
$c_i^* = g_i = 30$ against $g_j = 10$ *iff* $\beta_i > -0.6$
$c_i^* = g_i = 0$ against $g_j = 20$ *iff* $\beta_i < -0.3$
$c_i^* = g_i = 30$ against $g_j = 20$ *iff* $\beta_i > -0.3$
$c_i^* = g_i = 0$ against $g_j = 30$ *iff* $\beta_i < -0.2$
$c_i^* = g_i = 30$ against $g_j = 30$ *iff* $\beta_i > -0.2$

QED.

A.2.6. Social Efficiency preferences

A.2.6.1. Proof of proposition 5

Proposition 5. *If subject i maximizes the utility function* $U_i^{SE}\left(\pi_i(g_i, g_j), \pi_j(g_i, g_j)\right)$, *where i contributes* g_i *and the other player contributes* g_j, *then subject i's contribution attitudes, denoted as* c_i^*, *will be*

(i), in the Social Dilemma,

(a) $c_i^* = g_i = 0 \; \forall g_j \in A_j \; iff \; p_i < \frac{1-m}{m}$

(b) $c_i^* = g_i \in A_i \; \forall g_j \in A_j \; iff \; p_i = \frac{1-m}{m}$

(c) $c_i^* = g_i = 30 \; \forall g_j \in A_j \; iff \; p_i > \frac{1-m}{m}$

(ii), in the Common Interest Game,

$$(\forall \beta_i), c_i^* = g_i = 30 \; \forall g_j \in A_j$$

Proof.

Let's start by writing the utility function of person i for generic levels of contribution g_i and g_j:

$$U_i^{SE}\left(\pi_i(g_i, g_j), \pi_j(g_i, g_j)\right) = (1 - p_i) \times \pi_i(g_i, g_j) + p_i \times \left(\pi_i(g_i, g_j) + \pi_j(g_i, g_j)\right)$$

Expanding the RHS, we get:

$$U_i^{SE}\left(\pi_i(g_i, g_j), \pi_j(g_i, g_j)\right) = \pi_i(g_i, g_j) - p_i \times \pi_i(g_i, g_j) + p_i \times \pi_i(g_i, g_j) + p_i \times \pi_j(g_i, g_j)$$

Given that $-p_i \times \pi_i(g_i, g_j) + p_i \times \pi_i(g_i, g_j) = 0$ and simplifying, we get:

$$U_i^{SE}\left(\pi_i(g_i, g_j), \pi_j(g_i, g_j)\right) = \pi_i(g_i, g_j) + p_i \times \pi_j(g_i, g_j)$$

Substituting both $\pi_i(g_i, g_j)$ and $\pi_j(g_i, g_j)$ by the material payoff function defined in chapter 4, we get:

$$U_i^{SE}\left(\pi_i(g_i, g_j), \pi_j(g_i, g_j)\right) = 30 - g_i + m \times (g_i + g_j) + p_i \times \{30 - g_j + m \times (g_i + g_j)\}$$

Once we have expressed the utility of person i explicitly in terms of g_i and g_j, we can calculate the marginal utility with respect to g_i to see whether person i increases or decreases his or her utility in his or her own contributions:

$$\frac{\partial U_i^{SE}\left(\pi_i(g_i, g_j), \pi_j(g_i, g_j)\right)}{\partial g_i} = -1 + m + p_i \times m$$

Note that, whenever $\overline{m} \in (1, \infty)$, the marginal utility becomes:

$$\frac{\partial U_i^{SE}\left(\pi_i(g_i, g_j), \pi_j(g_i, g_j)\right)}{\partial g_i} = -1 + (> 1) \times (1 + p_i)$$

Given that $p_i \in [0,1]$, the marginal utility will always be positive, as:

$$\frac{\partial U_i^{SE}}{\partial g_i} = -1 + (> 1) \times \big(1 + (\geq 0)\big) = -1 + (> 1) \times (\geq 1) = -1 + (> 1) = (> 0)$$

Hence, the best response for a common interest game, where $\overline{m} \in (1, \infty)$ is given by:

$$(\forall\, p_i[0,1])\,, c_i^* = \; g_i = 30 \;\forall g_j \in A_j$$

Which proves (ii).

In a social dilemma, where $\underline{m} \in \left(\frac{1}{n}, 1\right)$, the value of the marginal utility can be positive or negative depending on the value of p_i. To find for which values of p_i does the marginal utility of $g_i$becomes negative, we just isolate p_i in the LHS of the marginal utility found above to get:

$$p_i \times \underline{m} < 1 - \underline{m}$$

Which, dividing both hand sides by $\underline{m}$, becomes:

$$p_i < \frac{1-\underline{m}}{\underline{m}}$$

Hence, when $p_i < \frac{1-\underline{m}}{\underline{m}}$ (resp. $p_i > \frac{1-\underline{m}}{\underline{m}}$) the utility of person i decreases (resp. increases) as he or she increases (resp. decreases) his or her contributions. Hence, the best response is given by:

$$c_i^* = \begin{cases} g_i = 0 \; \forall g_j \in A_j & if \; p_i < \frac{1-\underline{m}}{\underline{m}} \\ g_i \in A_i \; \forall g_j \in A_j & if \; p_i = \frac{1-\underline{m}}{\underline{m}} \\ g_i = 30 \; \forall g_j \in A_j & if \; p_i > \frac{1-\underline{m}}{\underline{m}} \end{cases}$$

Which proves all points in (i).

QED.

A.2.6.2. Other results involving social efficiency preferences

Below we provide a corollary that presents the specific threshold values of p_i determining optimal contributions for each g_j.

Corollary 5.1.: *If subject* i *maximizes the utility function* $U_i^{SE}\left(\pi_i(g_i, g_j), \pi_j(g_i, g_j)\right)$*, and given* $\underline{m} = 0.6$ *and* $\overline{m} = 1.2$*, the subject* i*'s choices will*

(i), in the Social Dilemma, be

$$c_i^* = \begin{cases} g_i = 0 \; \forall g_j \in A_j & if \; p_i < \frac{2}{3} \\ g_i \in A_i \; \forall g_j \in A_j & if \; p_i = \frac{2}{3} \\ g_i = 30 \; \forall g_j \in A_j & if \; p_i > \frac{2}{3} \end{cases}$$

(ii), in the Common Interest Game, be

$$(\forall p_i), g_i = 30 \; \forall \; g_j \in A_j$$

Proof.

(a) Given the best response for the social dilemma found in proposition 5, and substituting $\underline{m} = 0.6$, we get:

$$c_i^* = \begin{cases} g_i = 0\ \forall g_j \in A_j & if\ p_i < \frac{2}{3} \\ g_i \in A_i\ \forall g_j \in A_j & if\ p_i = \frac{2}{3} \\ g_i = 30\ \forall g_j \in A_j & if\ p_i > \frac{2}{3} \end{cases}$$

Which proves (i). Point (ii) is self-evident given proposition 5.

QED.

A.2.7. Maximin preferences

A.2.7.1. Proof of proposition 6

Proposition 6. *If subject i maximizes the utility function $U_i^{MM}\left(\pi_i(g_i, g_j), \pi_j(g_i, g_j)\right)$, where i contributes g_i and the other player contributes g_j, then subject i's contribution attitudes, denoted as c_i^*, will be*

(i), in the Social Dilemma,

(a) $c_i^* = g_i = 0\ \forall g_j \in A_j\ iff\ q_i < 1 - \underline{m}$

(b) $c_i^* = g_i \in [0, g_j]\ \forall g_j \in A_j\ iff\ q_i = 1 - \underline{m}$

(c) $c_i^* = g_i = g_j\ \forall g_j \in A_j\ iff\ q_i > 1 - \underline{m}$

(ii), in the Common Interest Game,

$$(\forall \beta_i), c_i^* = g_i = 30\ \forall g_j \in A_j$$

Proof.

Let's start by writing the utility function of person i for generic levels of contribution g_i and g_j:

$$U_i^{MM}\left(\pi_i(g_i,g_j),\pi_j(g_i,g_j)\right) = (1-q_i)\times\pi_i(g_i,g_j) + q_i\times min\{\pi_i(g_i,g_j),\pi_j(g_i,g_j)\}$$

Using the results of Lemma 0 (a), we know that $min\{\pi_i(g_i,g_j),\pi_j(g_i,g_j)\} = \pi_i(g_i,g_j)$ whenever $g_i > g_j$ and $min\{\pi_i(g_i,g_j),\pi_j(g_i,g_j)\} = \pi_j(g_i,g_j)$ whenever $g_i < g_j$. Hence, we can rewrite the previous utility function as follows:

$$U_i\left(\pi_i(g_i,g_j),\pi_j(g_i,g_j)\right) = \begin{cases}(1-q_i)\times\pi_i(g_i,g_j) + q_i\times\pi_i(g_i,g_j)\ if\ g_i \geq g_j\\(1-q_i)\times\pi_i(g_i,g_j) + q_i\times\pi_j(g_i,g_j)\ if\ g_i < g_j\end{cases}$$

By taking $\pi_i(g_i,g_j)$ as a common factor when $g_i \geq g_j$ and expanding the first parenthesis when $g_i < g_j$, we get:

$$U_i\left(\pi_i(g_i,g_j),\pi_j(g_i,g_j)\right) = \begin{cases}\pi_i(g_i,g_j)\times(1-q_i+q_i)\ if\ g_i \geq g_j\\\pi_i(g_i,g_j) - q_i\times\pi_i(g_i,g_j) + q_i\times\pi_j(g_i,g_j)\ if\ g_i < g_j\end{cases}$$

Simplifying when $g_i \geq g_j$ and taking q_i as a common factor when $g_i < g_j$, we get:

$$U_i\left(\pi_i(g_i,g_j),\pi_j(g_i,g_j)\right) = \begin{cases}\pi_i(g_i,g_j)\ if\ g_i \geq g_j\\\pi_i(g_i,g_j) + q_i\times\left(\pi_j(g_i,g_j) - \pi_i(g_i,g_j)\right)\ if\ g_i < g_j\end{cases}$$

Using Lemma 0 (b), we can substitute $\pi_j(g_i,g_j) - \pi_i(g_i,g_j) = g_i - g_j$ when $g_i < g_j$ to get:

$$U_i\left(\pi_i(g_i,g_j),\pi_j(g_i,g_j)\right) = \begin{cases}\pi_i(g_i,g_j) & if\ g_i \geq g_j\\\pi_i(g_i,g_j) + q_i\times(g_i - g_j)\ if\ g_i < g_j\end{cases}$$

Substituting $\pi_i(g_i,g_j)$ by the corresponding material payoff function outlined above, we get:

$$U_i\left(\pi_i(g_i, g_j), \pi_j(g_i, g_j)\right) = \begin{cases} 30 - g_i + m \times (g_i + g_j) & if\ g_i \geq g_j \\ 30 - g_i + m \times (g_i + g_j) + q_i \times (g_i - g_j) & if\ g_i < g_j \end{cases}$$

Taking the marginal derivative of person i's utility function with respect to his or her own contributions, we get:

$$\frac{\partial U_i\left(\pi_i(g_i, g_j), \pi_j(g_i, g_j)\right)}{\partial g_i} = \begin{cases} -1 + m\ if\ g_i \geq g_j \\ -1 + m + q_i\ if\ g_i < g_j \end{cases}$$

(a)

Note that, in a common interest game, where $\overline{m} \in (1, \infty)$, the marginal derivative of person i's utility function becomes positive regardless of the value of g_i. To see this, note that the first step takes the following values:

$$-1 + (> 1) = (> 0)$$

Given that $q_i \in [0,1]$, the second step takes the following values:

$$-1 + (> 1) + (\geq 0) = (> 0)$$

Hence, the optimal contribution for person i in the CIG becomes:

$$c_i^* = g_j = 30\ \forall g_j \in A_j\ \forall q_i \in [0,1]$$

Which proves (ii).

In a social dilemma game, where $\underline{m} \in \left(\frac{1}{n}, 1\right)$, the marginal derivative of person i's utility function becomes negative regardless of the value of q_i when $g_i \geq g_j$. This is so as $-1 + \underline{m}$ if always negative for $\underline{m} < 1$.

The marginal utility of own contribution when $g_i < g_j$ depends on the value of q_i. More specifically, the marginal utility will be positive in that range whenever the following inequality holds true:

$$-1 + \underline{m} + q_i > 0$$

Which implies the condition $q_i > 1 - \underline{m}$. Hence, when $q_i > 1 - \underline{m}$ a person will find it profitable to increase his contributions whenever $g_i < g_j$, and unprofitable to keep increasing his contributions in the range $g_i \geq g_j$. It, then, follows that the best response when $q_i > 1 - \underline{m}$ is to contribute $g_i = g_j$:

$$c_i^* = g_i = g_j \; \forall g_j \in A_j \; iff \; q_i > 1 - \underline{m}$$

Following an analogous logic, the best response when $q_i < 1 - \underline{m}$ is to contribute $g_i = 0$ for all g_j; as, subject to those parameter values, increasing contributions decreases utility in the range $g_i < g_j$. Hence,

$$c_i^* = g_i = 0 \; \forall g_j \in A_j \; iff \; q_i < 1 - \underline{m}$$

Finally, whenever $q_i = 1 - \underline{m}$, a person will be indifferent between any g_i in the range $[0, g_j]$, as the marginal utility does not vary with own contributions in this case.
More compactly, one can express those results as follows:

$$c_i^* = \begin{cases} g_i = 0 \; \forall g_j \in A_j & if \; q_i < 1 - \underline{m} \\ g_i \in [0, g_j] \; \forall g_j \in A_j & if \; q_i = 1 - \underline{m} \\ g_i = g_j \; \forall g_j \in A_j & if \; q_i > 1 - \underline{m} \end{cases}$$

Which proves (i).

QED.

A.2.7.2. Other results involving maximin preferences

Below we provide a corollary that presents the specific threshold values of q_i determining optimal contributions for each g_j.

Corollary 6.1. *If subject* i *maximizes the utility function* $U_i^{MM}\left(\pi_i(g_i,g_j),\pi_j(g_i,g_j)\right)$*, and given* $\underline{m} = 0.6$ *and* $\overline{m} = 1.2$*, the subject* i*'s choices will*

(i), in the Social Dilemma, be

$$c_i^* = \begin{cases} g_i = 0\ \forall g_j \in A_j & if\ q_i < 0.4 \\ g_i \in A_i\ \forall g_j \in A_j & if\ q_i = 0.4 \\ g_i = 30\ \forall g_j \in A_j & if\ q_i > 0.4 \end{cases}$$

(ii), in the Common Interest Game, be

$$(\forall q_i), g_i = 30\ \forall\ g_j \in \{0,10,20,30\}$$

Proof.

Given the best response for the social dilemma found in proposition 6, and substituting $\underline{m} = 0.6$, we get:

$$c_i^* = \begin{cases} g_i = 0\ \forall g_j \in A_j & if\ q_i < 0.4 \\ g_i \in A_i\ \forall g_j \in A_j & if\ q_i = 0.4 \\ g_i = 30\ \forall g_j \in A_j & if\ q_i > 0.4 \end{cases}$$

Which proves (i). Point (ii) is self-evident given proposition 6.

QED.

A.3. Proofs regarding estimated parameters through the use of parameter-elicitation games

A.3.1. Ultimatum Games

In the derivations below, we use the following notation:

- $x \in [0,7]$ represents the offer made by the sender
- 14 is the initial endowment of the sender
- 0 is the quantity that both get if the receiver rejects the sender's offer
- ε is an arbitrarily small number representing the smallest increase and or decrease of an offer.
- i is referred to as the receiver, and hence $U_i()$ represents the utility of the receiver
- A given distribution $(x, 14 - x)$ represents the payoff of the receiver in the first place (x) and the payoff of the sender in the second place ($14 - x$). That is, we define $\pi_i(x, 14 - x) = x$ and $\pi_j(x, 14 - x) = 14 - x$.

A.3.1.1. Disadvantageous Inequality parameter

A.3.1.1.1. Proof of proposition 7

Proposition 7. *If subject* i *maximizes the utility function* $U_i^{FS}\left(\pi_i(x, 14-x), \pi_j(x, 14-x)\right)$*, subject* i*'s minimum acceptable offer is* $x+\varepsilon+\varepsilon$ *and subject* i*'s maximum rejectable offer is* $x+\varepsilon$*, where* $x+\varepsilon+\varepsilon \leq 7$ *and* $x+\varepsilon \geq 0$*, then subject* i*'s choices would reveal an* α_i *parameter between the following boundaries:*

$$\frac{x+\varepsilon}{14-2\times(x+\varepsilon)} < \alpha_i < \frac{x+\varepsilon+\varepsilon}{14-2\times(x+\varepsilon+\varepsilon)}$$

Proof.

As a generic offer $x \in [0,7]$, then it follows that $14-x \in [7,14]$. Hence, $14-x \geq x$, and no offer goes above 7 regardless of the value of ε. This means that $U_i^{FS}\left(\pi_i(x, 14-x), \pi_j(x, 14-x)\right)$ will be on the domain of disadvantageous inequality as $14-x \geq x$ implies $\pi_i(x, 14-x) < \pi_j(x, 14-x)$. Hence, $U_i^{FS}\left(\pi_i(x, 14-x), \pi_j(x, 14-x)\right)$for the 2-person ultimatum game described above, for a generic offer x, is:

$$U_i^{FS}(x, 14-x) = 14-x-\alpha_i\times(14-x-x)$$

To compute the generic threshold of α_i, we assume a person's minimum acceptable offer is $x+\varepsilon+\varepsilon$ and his or her maximum rejectable offer is $x+\varepsilon$ as stated in the proposition, where $\varepsilon \geq 0$, and $x+\varepsilon+\varepsilon \leq 7$. This would imply that the utility of accepting the minimum acceptable offer is greater than the utility of the distribution $(0,0)$ and that the utility of accepting the maximum rejectable offer is lower than the utility of the distribution $(0,0)$. In mathematical terms:

$$U_i^{FS}(x+\varepsilon, 14-x-\varepsilon) < U_i^{FS}(0,0)$$
$$U_i^{FS}(x+\varepsilon+\varepsilon, 14-x-\varepsilon-\varepsilon) > U_i^{FS}(0,0)$$

Substituting the generic utility function by the Fehr-Schmidt specification presented in chapter 4, the two equations above would transform into:

$$x + \varepsilon - \alpha_i \times \big(14 - x - \varepsilon - (x + \varepsilon)\big) < 0$$

$$x + \varepsilon + \varepsilon - \alpha_i \times \big(14 - x - \varepsilon - \varepsilon - (x + \varepsilon + \varepsilon)\big) > 0$$

Simplifying, we get:

$$x + \varepsilon - \alpha_i \times \big(14 - 2 \times (x + \varepsilon)\big) < 0$$

$$x + \varepsilon + \varepsilon - \alpha_i \times \big(14 - 2 \times (x + \varepsilon + \varepsilon)\big) > 0$$

Which collapse to:

$$\alpha_i > \frac{(x + \varepsilon)}{14 - 2 \times (x + \varepsilon)}$$

$$\alpha_i < \frac{(x + \varepsilon + \varepsilon)}{14 - 2 \times (x + \varepsilon + \varepsilon)}$$

And, hence, it follows that:

$$\frac{(x + \varepsilon)}{14 - 2 \times (x + \varepsilon)} < \alpha_i < \frac{(x + \varepsilon + \varepsilon)}{14 - 2 \times (x + \varepsilon + \varepsilon)}$$

QED.

A.3.1.1.2. More proofs on the disadvantageous inequality parameter elicitation

As we showed in corollary 2.1 (b), the key value of the disadvantageous inequality parameter for our predictions of inequality aversion preferences regarding cooperation attitudes in the CIG is $\alpha_i \gtreqless 0.2$. Below we provide a corollary showing that a minimum acceptable offer (resp. maximum rejectable offer) of $x = 2$ precisely reveals this threshold.

Corollary 7.1. *Let's suppose that subject* i *maximizes the utility function* $U_i^{FS}\left(\pi_i(x, 14 - x), \pi_j(x, 14 - x)\right)$. *Then, if subject* i*'s minimum acceptable offer is 2 or lower subject* i *reveals* $\alpha_i < 0.2$. *If subject* i*'s maximum rejectable offer is 2 or higher subject* i *reveals* $\alpha_i > 0.2$

Given the inequalities found in Proposition 7, it follows that a minimum acceptable offer of 2 or lower would entail:

$$\alpha_i < \frac{(\leq 2)}{\left(14 - 2 \times \left((\leq 2)\right)\right)}$$

Similarly, a maximum rejectable offer of 2 or higher would entail:

$$\alpha_i > \frac{(\geq 2)}{\left(14 - 2 \times \left((\geq 2)\right)\right)}$$

And, hence,

$$\alpha_i < \frac{(\leq 2)}{\left(14 - (\leq 4)\right)}$$

$$\alpha_i > \frac{(\geq 2)}{\left(14 - (\geq 4)\right)}$$

Which becomes:

$$\alpha_i < \frac{(\leq 2)}{(\geq 10)}$$

$$\alpha_i > \frac{(\geq 2)}{(\leq 10)}$$

Now, let's define a partially ordered set:

$$P := (X, \leq)$$

Where

$$X \coloneqq \{x \in X | x \geq 0\}$$

We define the set MAO as follows:

$$MAO \coloneqq \left\{x \in MAO | \left((x \in X) \wedge \left(x < \frac{(\leq 2)}{(\geq 10)}\right)\right)\right\}$$

And, also, we define the set MRO as follows:

$$MRO \coloneqq \left\{x \in MRO | \left((x \in X) \wedge \left(x > \frac{(\geq 2)}{(\leq 10)}\right)\right)\right\}$$

Where MAO stands for '*Minimum Acceptable Offer*' and MRO stands for '*Maximum Rejectable Offer*'. It is straightforward to see that MAO is bounded above by $y = \frac{2}{10}$, as (i) $y \geq x \; \forall x \in MAO$ and (ii) $y \geq 0$ and, hence, $y \in X$.

Using a similar logic, it is also straightforward to see that MRO is bounded below by $y = \frac{2}{10}$, as (i) $x \geq y \; \forall x \in MRO$ and (ii) y$\geq$ 0 and, hence, $y \in X$.

Given that $y = \frac{2}{10}$ is an upper (lower) bound of MAO (MRO), and that it is the lowest upper bound (greatest lower bound) of MAO (MRO), it trivially follows that:

$$maxMAO = supMAO = \frac{2}{10} \in X$$

$$minMRO = infMRO = \frac{2}{10} \in X$$

It, then, follows, that the values of α_i for the first (second) inequality found above must be lower than the supremum of MAO (greater than the infimum of MRO):

$$\alpha_i < SupMAO$$

$$\alpha_i > InfMAO$$

And, substituting the values of $supMAO$ and $infMRO$, we get:

$$\alpha_i < 0.2$$

$$\alpha_i > 0.2$$

It follows that a person whose minimum acceptable offer is 2 or lower reveals $\alpha_i < 0.2$ and a person whose maximum rejectable offer is 2 or higher reveals $\alpha_i > 0.2$

QED.

A.3.2. Modified Dictator Games

In the derivations below, we use the following notation:

- $(20,0)$ is the original allocation that the dictator can choose instead of the equitable allocation

- $x \in [0,32]$ refers to the value that each gets from the equitable allocation. Hence, a given distribution (x, x) represents the payoff of the dictator and the receiver.

- ε is an arbitrarily small number representing the smallest increase and or decrease in the value each gets from the equitable allocation.

- i is referred to as the dictator, and hence $U_i()$ represents the utility of the dictator

A.3.2.1. Advantageous Inequality and Spiteful parameters

A.3.2.1.1. Proof of proposition 8

Proposition 8. *If subject* i *that maximizes the utility function* $U_i^{FS}(\pi_i, \pi_j)$, *whose maximum rejection quantity is* $x + \varepsilon$, *from the distribution* $(x + \varepsilon, x + \varepsilon)$, *to accept a distribution* $(20,0)$, *and whose minimum accepting quantity is* $x + \varepsilon + \varepsilon$, *from the distribution* $(x + \varepsilon + \varepsilon, x + \varepsilon + \varepsilon)$, *to reject a distribution* $(20,0)$, *will have a* β_i *parameter within the following boundaries:*

$$\frac{20 - (x + \varepsilon + \varepsilon)}{20} < \beta_i < \frac{20 - (x + \varepsilon)}{20}$$

Proof.

Let's assume a person with $U_i^{FS}(\pi_i, \pi_j)$ reveals the following preference pattern with their choices in the modified dictator games:

$$U_i^{FS}(20,0) > U_i^{FS}(x + \varepsilon, x + \varepsilon)$$

$$U_i^{FS}(20,0) < U_i^{FS}(x + \varepsilon + \varepsilon, x + \varepsilon + \varepsilon)$$

Substituting the generic utility by the inequality aversion preferences, the equations can be rewritten as:

$$20 - \beta_i \times (20) > x + \varepsilon$$

$$20 - \beta_i \times (20) < x + \varepsilon + \varepsilon$$

Isolating β_i in the RHS, we get:

$$20 - (x + \varepsilon) > \beta_i \times (20)$$

$$20 - (x + \varepsilon + \varepsilon) < \beta_i \times (20)$$

Which simplify to:

$$\frac{20 - (x + \varepsilon)}{20} > \beta_i$$

$$\frac{20 - (x + \varepsilon + \varepsilon)}{20} < \beta_i$$

Hence, β_i can be expressed in terms of the two thresholds together:

$$\frac{20 - (x + \varepsilon + \varepsilon)}{20} < \beta_i < \frac{20 - (x + \varepsilon)}{20}$$

QED.

A.3.2.1.2. More proofs on the advantageous inequality and spiteful parameters elicitation

As we showed in corollary 2.1 (a), the key value of the advantageous inequality parameter for our predictions of inequality aversion preferences regarding cooperation attitudes in the SDG is $\beta_i \gtreqless 0.4$. Also, corollary 4.1 showed that the relevant parameter values of β_i for play in the

CIG were $\beta_i \gtreqless -0.6$, $\beta_i \gtreqless -0.3$, and $\beta_i \gtreqless -0.2$. Below we provide a corollary showing that a maximum rejecting quantity (resp. minimum accepting quantity) of $x = 12$ reveals the necessary threshold for the inequality aversion model, and that a maximum rejecting quantity (resp. minimum accepting quantity) of $x = 24$, $x = 26$ and $x = 32$ reveal the necessary thresholds for predictions of cooperation attitudes in the CIG for the spiteful preferences model.

Corollary 8.1. *Let's suppose that subject* i *maximizes the utility function* $U_i^{FS}(\pi_i, \pi_j)$*. Then,*

(a) If subject i*'s minimum accepting quantity is 12 or lower subject* i *reveals* $\beta_i > 0.4$*. If subject* i*'s maximum rejecting quantity is 12 or higher subject* i *reveals* $\beta_i < 0.4$.

(b) If subject i*'s minimum accepting quantity is 24 or lower subject* i *reveals* $\beta_i > -0.2$*. If subject* i*'s maximum rejecting quantity is 24 or higher subject* i *reveals* $\beta_i < -0.2$.

(c) If subject i*'s minimum accepting quantity is 26 or lower subject* i *reveals* $\beta_i > -0.3$*. If subject* i*'s maximum rejecting quantity is 26 or higher subject* i *reveals* $\beta_i < -0.3$.

(d) If subject i*'s minimum accepting quantity is 32 or lower subject* i *reveals* $\beta_i > -0.6$*. . If subject* i*'s maximum rejecting quantity is 32 or higher subject* i *reveals* $\beta_i < -0.6$.

Proof.

(a)

Given the inequality found in Proposition 8, it follows that a minimum accepting quantity of 12 or lower would entail:

$$\beta_i > \frac{20 - (\leq 12)}{20}$$

Similarly, a maximum rejecting quantity of 2 or higher would entail:

$$\beta_i < \frac{20 - (\geq 12)}{20}$$

And, hence,

$$\beta_i > \frac{\geq 8}{20}$$

$$\beta_i < \frac{\leq 8}{20}$$

Now, let's define a partially ordered set:

$$P := (X, \leq)$$

Where

$$X \coloneqq \{x \in X | x \geq 0\}$$

We define the set MAQ as follows:

$$MAQ \coloneqq \left\{x \in MAQ | \left((x \in X) \wedge \left(x > \frac{\geq 8}{20}\right)\right)\right\}$$

And, also, we define the set MRO as follows:

$$MRQ \coloneqq \left\{x \in MRQ | \left((x \in X) \wedge \left(x < \frac{\leq 8}{20}\right)\right)\right\}$$

Where MAQ stands for '*Minimum Accepting Quantity*' and MRO stands for '*Maximum Rejecting Quantity*'. It is straightforward to see that MAQ is bounded below by $y = \frac{8}{20}$, as (i) $y \leq x \; \forall x \in MAQ$ and (ii) $y \geq 0$ and, hence, $y \in X$.

Using a similar logic, it is also straightforward to see that MRQ is bounded above by $y = \frac{8}{20}$, as (i) $y \geq x \; \forall x \in MRQ$ and (ii) y≥ 0 and, hence, $y \in X$.

Given that $y = \frac{8}{20}$ is a lower (upper) bound of MAQ (MRQ), and that it is the greatest lower bound (least upper bound) of MAQ (MRQ), it trivially follows that:

$$minMAQ = infMAQ = \frac{8}{20} \in X$$

$$maxMRQ = supMRQ = \frac{8}{20} \in X$$

It, then, follows, that the values of β_i for the first (second) inequality found above must be greater than the infimum of MAQ (lower than the supremum of MRQ):

$$\beta_i > infMAQ$$

$$\beta_i < supMRQ$$

And, substituting the values of $infMAQ$ and $supMRQ$, we get:

$$\beta_i > 0.4$$

$$\beta_i < 0.4$$

It follows that a person whose minimum accepting quantity is 12 or lower reveals $\beta_i > 0.4$ and a person whose maximum rejecting quantity is 12 or higher reveals $\beta_i < 0.4$

(b)

Following (a), a minimum accepting quantity of 24 or lower and a maximum rejecting quantity of 24 or higher would entail:

$$\beta_i > \frac{20 - (\leq 24)}{20}$$

$$\beta_i < \frac{20 - (\geq 24)}{20}$$

And, hence,

$$\beta_i > \frac{-(\leq 4)}{20}$$

$$\beta_i < \frac{-(\geq 4)}{20}$$

Now, let's define a partially ordered set:

$$P := (X, \leq)$$

Where

$$X \coloneqq \{x \in X | x \leq 0\}$$

We define the set MAQ as follows:

$$MAQ := \left\{x \in MAQ | \left((x \in X) \wedge \left(x > \frac{-(\leq 4)}{20} \right) \right) \right\}$$

And, also, we define the set MRO as follows:

$$MRQ := \left\{x \in MRQ | \left((x \in X) \wedge \left(x < \frac{-(\geq 4)}{20} \right) \right) \right\}$$

It is straightforward to see that MAQ is bounded below by $y = -\frac{4}{20}$, as (i) $y \leq x \; \forall x \in MAQ$ and (ii) $y \leq 0$ and, hence, $y \in X$.

Using a similar logic, it is also straightforward to see that MRQ is bounded above by $y = \frac{4}{20}$, as (i) $y \geq x \; \forall x \in MRQ$ and (ii) y$\leq$ 0 and, hence, $y \in X$.

Given that $y = -\frac{4}{20}$ is a lower (upper) bound of MAQ (MRQ), and that it is the greatest lower bound (least upper bound) of MAQ (MRQ), it trivially follows that:

$$minMAQ = infMAQ = -\frac{4}{20} \in X$$

$$maxMRQ = supMRQ = -\frac{4}{20} \in X$$

It, then, follows, that the values of β_i for the first (second) inequality found above must be greater than the infimum of MAQ (lower than the supremum of MRQ):

$$\beta_i > infMAQ$$

$$\beta_i < supMRQ$$

And, substituting the values of $infMAQ$ and $supMRQ$, we get:

$$\beta_i > -0.2$$

$$\beta_i < -0.2$$

It follows that a person whose minimum accepting quantity is 24 or lower reveals $\beta_i > -0.2$ and a person whose maximum rejecting quantity is 24 or higher reveals $\beta_i < -0.2$

(c)

Following (b), a minimum accepting quantity of 26 or lower and a maximum rejecting quantity of 26 or higher would entail:

$$\beta_i > \frac{20 - (\leq 26)}{20}$$

$$\beta_i < \frac{20 - (\geq 26)}{20}$$

And, hence,

$$\beta_i > \frac{-(\leq 6)}{20}$$

$$\beta_i < \frac{-(\geq 6)}{20}$$

Using the same technique as in (b), which we omit to avoid unnecessary repetition, it follows that:

$$\beta_i > -0.3$$

$$\beta_i < -0.3$$

It follows that a person whose minimum accepting quantity is 26 or lower reveals $\beta_i > -0.3$ and a person whose maximum rejecting quantity is 26 or higher reveals $\beta_i < -0.3$

(d)

Following (b), a minimum accepting quantity of 32 or lower and a maximum rejecting quantity of 32 or higher would entail:

$$\beta_i > \frac{20 - (\leq 32)}{20}$$

$$\beta_i < \frac{20 - (\geq 32)}{20}$$

And, hence,

$$\beta_i > \frac{-(\leq 12)}{20}$$

$$\beta_i < \frac{-(\geq 12)}{20}$$

Using the same technique as in (b), which we omit to avoid unnecessary repetition, it follows that:

$$\beta_i > -0.6$$

$$\beta_i < -0.6$$

It follows that a person whose minimum accepting quantity is 32 or lower reveals $\beta_i > -0.6$ and a person whose maximum rejecting quantity is 32 or higher reveals $\beta_i < -0.6$

QED.

A.3.2.2. Social Efficiency parameter

A.3.2.2.1. Proof of proposition 9.

Proposition 9. *Let's suppose that subject* i *maximizes the utility function* $U_i^{SE}(\pi_i, \pi_j)$. *If subject* i*'s maximum rejection quantity is* $x+\varepsilon$, *from the distribution* $(x+\varepsilon, x+\varepsilon)$, *to accept a distribution* $(20{,}0)$, *and if subject* i*'s minimum accepting quantity is* $x+\varepsilon+\varepsilon$, *from the distribution* $(x+\varepsilon+\varepsilon, x+\varepsilon+\varepsilon)$, *to reject a distribution* $(20{,}0)$, *then subject* i *will reveal to have a* p_i *parameter within the following boundaries:*

$$\frac{20-(x+\varepsilon+\varepsilon)}{x+\varepsilon+\varepsilon} < p_i < \frac{20-(x+\varepsilon)}{x+\varepsilon}$$

Proof.

Let's assume a person with $U_i^{SE}\left(\pi_i(x, 14-x), \pi_j(x, 14-x)\right)$ preferences reveals the following preference pattern with their choices in the modified dictator games:

$$U_i^{SE}(20{,}0) > U_i^{SE}(x+\varepsilon, x+\varepsilon)$$

$$U_i^{SE}(20{,}0) < U_i^{SE}(x+\varepsilon+\varepsilon, x+\varepsilon+\varepsilon)$$

These equations can be rewritten as:

$$(1-p_i)\times 20 + p_i \times (20) > (1-p_i)\times(x+\varepsilon) + p_i \times (x+\varepsilon+x+\varepsilon)$$

$$(1-p_i)\times 20 + p_i \times (20) < (1-p_i)\times(x+\varepsilon+\varepsilon) + p_i \times (x+\varepsilon+\varepsilon+x+\varepsilon+\varepsilon)$$

Which, by taking 20 as a common factor in the LHS and simplifying, can be rewritten as:

$$20 > x + \varepsilon - p_i \times (x + \varepsilon) + 2p_i \times (x + \varepsilon)$$

$$20 < x + \varepsilon + \varepsilon - p_i \times (x + \varepsilon + \varepsilon) + 2p_i \times (x + \varepsilon + \varepsilon)$$

Simplifying, we get:

$$20 > x + \varepsilon + p_i \times (x + \varepsilon)$$

$$20 < x + \varepsilon + \varepsilon + p_i \times (x + \varepsilon + \varepsilon)$$

Isolating p_i in the RHS, we get:

$$20 - (x + \varepsilon) > p_i \times (x + \varepsilon\ )$$

$$20 - (x + \varepsilon + \varepsilon) < p_i \times (x + \varepsilon + \varepsilon)$$

Which can be rewritten as:

$$\frac{20 - (x + \varepsilon)}{x + \varepsilon} > p_i$$

$$\frac{20 - (x + \varepsilon + \varepsilon)}{x + \varepsilon + \varepsilon} < p_i$$

Hence, p_i can be said to lie between the following boundaries:

$$\frac{20 - (x + \varepsilon + \varepsilon)}{x + \varepsilon + \varepsilon} < p_i < \frac{20 - (x + \varepsilon)}{x + \varepsilon}$$

QED.

A.3..2.2.2. More proofs on the social efficiency parameter elicitation

As we showed in corollary 5.1, the key value of the social efficiency parameter for our predictions of social efficiency preferences regarding cooperation attitudes in the SDG is $p_i \gtreqless \frac{2}{3}$. Below we provide a corollary showing that a maximum rejecting quantity (resp. minimum accepting quantity) of $x = 12$ reveals the necessary threshold for the social efficiency model to make predictions regarding play in the SDG.

Corollary 9.1. *Let's suppose that subject* i *maximizes the utility function* $U_i^{SE}(\pi_i, \pi_j)$*. Then,* if subject i's minimum accepting quantity is 12 or lower subject i reveals $p_i > \frac{2}{3}$. If subject i's maximum rejecting quantity is 12 or higher subject i reveals $p_i < \frac{2}{3}$

Given the inequality found in Proposition 9., it follows that a minimum accepting quantity of 12 or lower would entail:

$$p_i > \frac{20 - (\leq 12)}{(\leq 12)}$$

Similarly, a maximum rejecting quantity of 2 or higher would entail:

$$p_i < \frac{20 - (\geq 12)}{(\geq 12)}$$

And, hence,

$$p_i > \frac{\geq 8}{(\leq 12)}$$

$$p_i < \frac{\leq 8}{(\geq 12)}$$

Now, let's define a partially ordered set:

$$P := (X, \leq)$$

Where

$$X \coloneqq \{x \in X | x \geq 0\}$$

We define the set MAQ as follows:

$$MAQ \coloneqq \left\{x \in MAQ | \left((x \in X) \wedge \left(x > \frac{\geq 8}{(\leq 12)}\right)\right)\right\}$$

And, also, we define the set MRO as follows:

$$MRQ \coloneqq \left\{x \in MRQ | \left((x \in X) \wedge \left(x < \frac{\leq 8}{(\geq 12)}\right)\right)\right\}$$

Using the same techniques as in in the previous corollaries., it is straightforward to see that $y = \frac{8}{12}$ is a lower bound of MAQ and an upper bound of MRQ. Hence, it follows that:

$$p > \frac{2}{3}$$

$$p < \frac{2}{3}$$

It follows that a person whose minimum accepting quantity is 12 or lower reveals $p_i > \frac{2}{3}$ and a person whose maximum rejecting quantity is 12 or higher reveals $p_i < \frac{2}{3}$.

QED.

A.3.2.2.3. Maximim parameter

A.3.2.2.3.1. Proof of proposition 10

Proposition 10. *Let's suppose that subject* i *maximizes the utility function* $U_i^{MM}(\pi_i, \pi_j)$.

(a) If subject i*'s maximum rejection quantity is* $x + \varepsilon$*, from the distribution* $(x + \varepsilon, x + \varepsilon)$*, to accept a distribution* $(20,0)$*, and if subject* i*'s minimum accepting quantity is* $x + \varepsilon + \varepsilon$*, from the distribution* $(x + \varepsilon + \varepsilon, x + \varepsilon + \varepsilon)$*, to reject a distribution* $(20,0)$*, then subject* i *will reveal to have a* q_i *parameter within the following boundaries:*

$$\frac{20 - (x + \varepsilon + \varepsilon)}{x + \varepsilon + \varepsilon} < q_i < \frac{20 - (x + \varepsilon)}{x + \varepsilon}$$

(b) If subject i*'s maximum rejection quantity is* $x + \varepsilon$*, from the distribution* $(x + \varepsilon, x + \varepsilon)$*, to accept a distribution* $(20,0)$*, and if subject* i*'s minimum accepting quantity is* $(x + \varepsilon + \varepsilon, x + \varepsilon + \varepsilon)$*, to reject a distribution* $(20,0)$*, then subject* i *reveals a maximin parameter* q_i *within the same threshold of values as the advantageous inequality parameter* β_i.

Proof.

(a)

Let's assume a person with a utility $U_i^{MM}(\pi_i, \pi_j)$ reveals the following preference pattern with their choices in the modified dictator games:

$$U_i^{MM}(20,0) > U_i^{MM}(x + \varepsilon, x + \varepsilon)$$

$$U_i^{MM}(20,0) < U_i^{MM}(x + \varepsilon + \varepsilon, x + \varepsilon + \varepsilon)$$

These equations can be rewritten as:

$$(1 - q_i) \times 20 + q_i \times (0) > x + \varepsilon$$

$$(1 - q_i) \times 20 + q_i \times (0) < x + \varepsilon + \varepsilon$$

Expanding the parenthesis, we get:

$$20 - q_i 20 > x + \varepsilon$$

$$20 - q_i 20 < x + \varepsilon + \varepsilon$$

Isolating p in the RHS, we get:

$$20 - (x + \varepsilon) > q_i \times 20$$

$$20 - (x + \varepsilon + \varepsilon) < q_i \times 20$$

Which can be rewritten as:

$$\frac{20 - (x + \varepsilon)}{20} > q_i$$

$$\frac{20 - (x + \varepsilon + \varepsilon)}{20} < q_i$$

Hence, q_i lies within the following boundaries:

$$\frac{20 - (x + \varepsilon + \varepsilon)}{20} < p < \frac{20 - (x + \varepsilon)}{20}$$

Which proves (a).

(b)

Recall the boundaries of β_i as found on proposition 8.:

$$\frac{20-(x+\varepsilon+\varepsilon)}{20}<\beta_i<\frac{20-(x+\varepsilon)}{20}$$

And recall the boundaries of q_i found in (a):

$$\frac{20-(x+\varepsilon+\varepsilon)}{20}<q_i<\frac{20-(x+\varepsilon)}{20}$$

Therefore, it follows that, given the generic maximum rejection quantity $x+\varepsilon$ and the minimum accepting quantity $x+\varepsilon+\varepsilon$, the boundaries of the maximin parameter q_i and of the advantageous inequality β_i will be the same, which proves (b).

QED.

A.3.2.2.3.2. More proofs on the maximin parameter elicitation

As we showed in corollary 6.1, the key value of the maximin parameter for our predictions of maximin preferences regarding cooperation attitudes in the SDG is $q_i \gtreqless 0.4$. Below we provide a corollary showing that a maximum rejecting quantity (resp. minimum accepting quantity) of $x=12$ reveals the necessary threshold for the maximin model to make predictions regarding play in the SDG.

Corollary 10.1. *Let's suppose that subject i maximizes the utility function $U_i^{MM}(\pi_i,\pi_j)$. If subject i's minimum accepting quantity is 12 or lower, then subject i reveals $q_i>0.4$. If subject i's maximum rejecting quantity is 12 or higher, then subject i reveals $p<0.4$*

Proof.

Given that proposition 10. (b) shows that the values of β_i and q_i coincide for generic maximum rejection and minimum accepting quantities, this proof is identical to that of Corollary 8.1. (a) and, hence, has already been proven.

QED.

A.3.3. Reciprocity Games

We use a modified version of the reciprocity games used in Bruhin et al (2019) to elicit the $Y_{i,j}$ parameter values of the Dufwenberg and Kirchsteiger utility function outlined in chapter 4. We impose certain restrictions on the values of each of the three allocations strategically to simplify the finding on the threshold values for $Y_{i,j}$. More specifically, the allocations are such that some strategies are inefficient in Dufwenberg and Kirchsteiger's (2004) model, thereby simplifying the calculations. The paragraph below summarises our specific setting of the reciprocity games we present to subjects:

Person j could choose $a_j = E$, which will enforce the distribution (x_1, x_5), or alternatively could choose $a_j = N$, which would give person i the possibility to choose between $a_i = A$, generating a distribution of (x_2, x_4) and $a_i = B$, generating a distribution of (x_3, x_6), where $x_1 > x_2 > x_3$ and $x_4 > x_5 > x_6$.

It is important to note before proceeding that, given the Dufwenberg and Kirchsteiger (2004) model we use, the restrictions on the values we impose on x_1, x_2, x_3, x_4, x_5 and x_6 imply the following:

a) Strategy $a_i = B$ is inefficient, as $x_2 > x_3$ and $x_4 > x_6$, and hence both players would be better off by playing $a_i = B$.
b) Strategy $a_j = N$ is not inefficient. Whereas it is true that for one subsequent history of play (namely, $a_i = B$) both players end worse off by player j having played $a_j = N$, as $x_3 < x_1$ and $x_6 < x_5$, for at least another subsequent history of play ($namely, a_i = A$) at least one player is better off by player j having played $a_j = N$, as $x_4 > x_5$ even when $x_2 < x_1$.

A.3.3.1. Reciprocity parameter – Proof of proposition 11.

Proposition 11. *Let's suppose that subject* i *maximizes the utility function* $U_i^{DK}(\pi_i, \pi_j)$. *Then,* *(a) Assuming beliefs are in equilibrium, a player i's choice of $a_i = A$ over $a_i = B$ given that the first mover has done $a_j = N$ implies the following about the reciprocity parameter:*

$$Y_{i,j} < \frac{2 \times (x_2 - x_3)}{(x_4 - x_6) \times (x_1 - x_2)}$$

(b) Assuming beliefs are in equilibrium, a player i's choice of $a_i = B$ over $a_i = A$ given that the first mover has done $a_j = N$ implies the following about the reciprocity parameter:

$$Y_{i,j} > \frac{2 \times (x_2 - x_3)}{(x_4 - x_6) \times (x_1 - x_3)}$$

Proof.

Given that the first mover has done $a_j = N$, the first-order belief of player i is updated so that $b_{ij}(h) = N$. The kindness function of player i towards player j reads:

$$\begin{aligned}\kappa_i\big(a_{ij}(h), b_{ij}(h) = N\big) \\ &= \pi_j\big(a_{ij}(h), b_{ij}(h) = N\big) \\ &- \frac{max\, \pi_j\big(a_{ij}(h), N\big) | a_i \in A_i + min\, \pi_j\big(a_{ij}(h), N\big) | a_i \in E_i}{2}\end{aligned}$$

Hence, given that only $a_i = A$ is the only efficient strategy for player i as discussed above, it follows that:

$$\kappa_i\big(a_{ij}(h) = A, b_{ij}(h) = N\big) = x_4 - x_4 = 0$$

$$\kappa_i\big(a_{ij}(h) = B, b_{ij}(h) = N\big) = x_6 - x_4 = -(x_4 - x_6)$$

To find the perceived kindness function, note that $(p'', A; 1 - p'', B)$ is the probability distribution for the second-order belief of person i. Hence, we can write the perceived kindness function as:

$$\lambda_{iji}\left(b_{ij}(h)=N,c_{iji}(h)\right)=\pi_i\left(b_{ij}(h)=N,c_{iji}(h)\right)-\frac{x_1+p''\times x_2+(1-p'')\times x_3}{2}.$$

Using $(p'',A;1-p'',B)$ to compute the expected payoff that player j intends to give player i by doing $b_{ij}(h)=N$, we get:

$$\lambda_{iji}\left(b_{ij}(h)=N,c_{iji}(h)\right)=p''\times\pi_i\big(b_{ij}(h)=N,a_i=A\big)+(1-p'')\times\pi_i\big(b_{ij}(h)=N,a_i=B\big)-\frac{x_1+p''\times x_2+(1-p'')\times x_3}{2}.$$

Which, after substituting the relevant payoffs, becomes:

$$\lambda_{iji}\left(b_{ij}(h)=N,c_{iji}(h)\right)=p''\times x_2+(1-p'')\times x_3-\frac{x_1+p''\times x_2+(1-p'')\times x_3}{2}$$

Rearranging, we get:

$$\lambda_{iji}\left(b_{ij}(h)=N,c_{iji}(h)\right)=p''\times x_2+(1-p'')\times x_3-\frac{x_1}{2}-\frac{p''\times x_2+(1-p'')\times x_3}{2}$$

Taking $p''\times x_2+(1-p'')\times x_3$ as a common factor and simplifying, we get:

$$\lambda_{iji}\left(b_{ij}(h)=N,c_{iji}(h)\right)=-\frac{x_1}{2}+\frac{p''\times x_2+(1-p'')\times x_3}{2}<0$$

Given the perceived kindness that i believes j is displaying towards him, and the kindness of each possible action that i can do, we can write person i's utility of both actions as:

$$U_i\left(a_i(h)=A,b_{ij}(h)=N,c_{iji}(h)\right)=x_2+Y_{i,j}\times(0)\times\left(-\frac{x_1}{2}+\frac{p''\times x_2+(1-p'')\times x_3}{2}\right)$$
$$=x_2$$

$$U_i\left(a_i(h) = B, b_{ij}(h) = N, c_{iji}(h)\right)$$
$$= x_3 - Y_{i,j} \times (x_4 - x_6) \times \left(-\frac{x_1}{2} + \frac{p'' \times x_2 + (1 - p'') \times x_3}{2}\right)$$

(a)

For person i to choose the allocation which gives him the highest payoff ($a_i = A$) the following condition needs to hold:

$$U_i\left(a_i(h) = A, b_{ij}(h) = N, c_{iji}(h)\right) > U_i\left(a_i(h) = B, b_{ij}(h) = N, c_{iji}(h)\right)$$

Which is equivalent to the following expression:

$$x_2 > x_3 + Y_{ij} \times (x_4 - x_6) \times \left(\frac{x_1}{2} - \frac{p'' \times x_2 + (1 - p'') \times x_3}{2}\right)$$

Isolating $Y_{i,j}$ in the RHS, the previous expression becomes:

$$x_2 - x_3 > Y_{ij} \times (x_4 - x_6) \times \left(\frac{x_1}{2} - \frac{p'' \times x_2 + (1 - p'') \times x_3}{2}\right)$$

Dividing both sides of the inequality by $\left((x_4 - x_6) \times \left(\frac{x_1}{2} - \frac{p'' \times x_2 + (1-p'') \times x_3}{2}\right)\right)$, we get:

$$Y_{i,j} < \frac{(x_2 - x_3)}{(x_4 - x_6) \times \left(\frac{x_1}{2} - \frac{p'' \times x_2 + (1 - p'') \times x_3}{2}\right)}$$

Let's assume that second-order beliefs are in equilibrium. That is to say, if $U_i\left(a_i(h) = A, b_{ij}(h) = N, c_{iji}(h)\right) > U_i\left(a_i(h) = B, b_{ij}(h) = N, c_{iji}(h)\right)$ then the second-order belief that Person i has is that Person j believes that he'll player $a_i(h) = A$ with certainty. Hence, $p'' = 1$. This would, in turn, give us the following threshold:

If $U_i\left(a_i(h) = A, b_{ij}(h) = N, c_{iji}(h)\right) > U_i\left(a_i(h) = B, b_{ij}(h) = N, c_{iji}(h)\right)$ and, hence, $p'' = 1$, then by substituting $p'' = 1$ in the inequality above, we get:

$$Y_{i,j} < \frac{2 \times (x_2 - x_3)}{(x_4 - x_6) \times (x_1 - x_2)}$$

(b)

If the beliefs are in equilibrium, it also follows that, if $U_i\left(a_i(h) = A, b_{ij}(h) = N, c_{iji}(h)\right) < U_i\left(a_i(h) = B, b_{ij}(h) = N, c_{iji}(h)\right)$, then the second-order belief that Person i has is that Person *j* believes that he'll play $a_i(h) = B$ with certainty. Hence, $p'' = 0$.

If $U_i\left(a_i(h) = A, b_{ij}(h) = N, c_{iji}(h)\right) < U_i\left(a_i(h) = B, b_{ij}(h) = N, c_{iji}(h)\right)$ and, hence, $p'' = 0$, then by substituting $p'' = 0$ in the inequality above, we get:

$$Y_{i,j} > \frac{2 \times (x_2 - x_3)}{(x_4 - x_6) \times (x_1 - x_3)}$$

QED.