Inequity Aversion Preferences in the Dynamic Public Goods Game*

Wei Hu †

Toulouse School of Economics

June 10, 2011

Abstract

This paper studies the effect of inequity aversion (Fehr and Schmidt, 1999) in a “repeated” public goods game. We assume that agents care about the differences in their aggregate payoffs throughout the game, so that the repeated game is in fact a dynamic game. Inequity-averse agents then cooperate on equal and nearly efficient levels of aggregate contributions to the public goods. We introduce the concept of Limited Punishment for Deviations to alleviate a severe multiple equilibria problem. Importantly, selfish agents also actively contribute in this game. The results are also robust to asymmetric information about agents’ types. Finally, the model predicts the end-game effect and the restart effect that are often observed in experimental games.

Keywords: Cooperation, Public Goods Game, Inequity Aversion

JEL codes: C73, D03, H41

*I am grateful to my supervisors Nicolas Treich and Thomas Mariotti. They have given invaluable suggestions when guiding my research. I also thank Giuseppe Attanasi, Cécile Aubert, Guillaume Cheikbossian, Jacques Crémer, Jean Tirole, and attendants at the PET10 conference in Istanbul, 2010 for many helpful comments.

†Contact: wei.hu@tse-fr.eu.
1 Introduction

How did cooperative behavior evolve? This question is on Science magazine’s list of top 25 “scientific puzzles that are driving basic scientific research”. For economists, cooperation does not come out naturally from the *Homo economicus* assumption. Non-selfish behavior certainly provides an explanation, but also raises several other questions. Specifically, how can such non-selfish behavior emerge, and is it sustainable when both selfish and non-selfish people interact? The objective of this paper is to shed some light on these questions using game theoretical tools.

We suppose that non-selfish behavior is driven by agents’ aversion to inequity (Fehr and Schmidt’s, 1999) and study it in a repeatedly played linear public goods game. An inequity-averse agent bears psychological costs on income differences between herself and others. The consequences of inequity aversion have been studied in different contexts including incentives in the workplace (e.g., Demougin and Fluet, 2003, Itoh, 2004, Fehr et al., 2007, Torgler et al., 2008), bargaining under competition (Fischbacher, Fong and Fehr, 2009), formation of long-term trading relationships (Brown, Falk and Fehr, 2004), collusion over product quantities (Santos-Pionto, 2006, Muller, 2006) and price rigidities of consumption goods (Rotemberg, 2004). But as far as we know, the current paper is the first to apply this model in a dynamic game.

For such an application, we first need to specify how inequity-averse agents care about income differences in a dynamic framework. We assume that they care about the differences in aggregate payoffs throughout the game. This means an agent’s strategy involves measuring and comparing expected payoffs in each round. If an inequity-averse agent cares enough about inequity, she will act to offset payoff differences already generated, and to prevent them from arising in the future. For instance, if another agent has previously contributed

---


2 This is a game in which agents decide how to invest their endowment in a public goods and a private goods. In labs it is usually carried out with 10 repetitions. Over-contribution and other patterns are commonly observed. We will talk more about it in the related literature part at the end of introduction. For a survey on public goods game studies, see Ledyard (1995).
more than her, she is inclined to contribute more in the future to reduce inequity.

Under this assumption, the “repeated” public goods game that we study here is in fact a dynamic game, where rounds are distinguished from one another by the accumulated payoff difference. The final round, for example, cannot be regarded the same as a static game. We therefore make different predictions from those obtained under a repeated game assumption. The main results of our approach are described in turn.

First, in a group of two inequity-averse agents, aggregate contributions are equal between individuals. In contrast to a repeated game, the contribution is bounded from below, and is close to efficiency. Any realization below this level, such as a zero contribution, is prone to deviations. The willingness of either agent to match contributions ensures that such deviations are profitable. There are, however, vast possibilities for behavior in each individual round. In an attempt to solve this multiple equilibria problem\(^3\), we introduce the principle of Limited Punishment for Deviations (LPD). Under LPD, the prediction for individual round’s contributions is unique and maximal for almost the entire game.

This high level cooperation is maintained even when selfish agents participate in the game. On top of that, selfish agents are proactive in cooperation formation. To understand this, imagine a group with one selfish and one inequity-averse agent. Suppose that contributions are set to zero. If the selfish agent starts to contribute before the game ends, the inequity-averse agent will match with the same amount in a later round. The return to the selfish’s move is then positive under the design of a public goods game. The existence of an inequity-averse agent therefore provides the incentive for the selfish one to cooperate. In the equilibrium where LPD is applied, a selfish agent will even make maximal contributions for all rounds but the final one. Similar cooperative predictions are also reached when informational asymmetry is introduced to agents’ types. In the extreme case where the prior belief about an

\(^3\)In the presence of conditionally cooperative agents, multiple equilibria are common. For example, Fehr and Schmidt (1999) show in their Proposition 4 that when there are contributions in a static public goods game, any feasible level can be symmetrically realized in the equilibrium.
agent being inequity-averse is low, we obtain the cooperative result by using the reputation game method of Kreps et al. (1982).

Although the current model does not predict a declining pattern in contribution across rounds, it does capture several regularities consistent with experimental evidences. In particular, the model predicts an end-game effect, which is a sudden drop of contributions by individual agents in the final rounds. This is due to the revealing of the selfish agents before the end of game. In contrast to what Fehr and Schmidt (1999) predict based on their calibrated model and the repeated game assumption, the final round’s contributions in our game remain positive. We also predict a restart effect that corresponds to the recovery of contributions in a surprise restart following the initially announced final round. This is caused by cooperation under complete information when at least one inequity-averse agent is present. The model also reveals that the inequity-averse agents may employ a “wait-and-see” strategy when playing the reputation game. By holding contribution at certain time, they enjoy the flexibility to catch up with others ex post and to avoid any payoff inequity. In this case, the selfish agents actively maintain public goods contributions in the game.

Related Literature

The current paper draws on the economics literature in game theory, behavioral economics, and experimental economics. We examine each of these areas of study in turn.

In the game theoretical literature, the folk theorem provides an explanation for cooperation in infinitely repeated games. The folk theorem says that cooperation can be sustained in equilibrium if players commit to punish deviators long and harsh enough, which eliminates all deviation benefits. Theoretically, any outcome that gives each player more than his/her minimax payoff can be supported as an equilibrium (Rubinstein, 1979). The cooperative equilibrium is just one of them, and there is however no prior reason for players to choose it. The folk theorem also applies to finite games where there is informational asymmetry concerning whether a player is a “committed” type (e.g., Fudenberg and Maskin, 1986, Schmidt, 1993). Results are then devel-
oped based on the reputation game demonstrated by Kreps and Wilson (1982) and Kreps et al. (1982). In that case, a rational agent mimics the behavior of a committed type in each round to benefit from building a reputation of being irrational. We borrow this idea in the current paper when informational asymmetry is present.

The studies of other-regarding preferences or social preferences have made a significant field in the literature of behavioral economics. These studies look at individuals’ concerns beyond own material payoffs. A typical example is the altruism preference, which describes one’s determined positive attitude towards others. Altruism has been studied in various subjects including family economics (Becker, 1974, Bernheim and Stark, 1988, Bergstrom, 1989), public goods (Haltiwanger and Waldman, 1992) and incentive theory (Ma, 2007). A nonlinear altruism model is shown to be also able to rationalize certain behavior in the lab (Andreoni and Miller, 2002). The interactive mechanism neglected in the altruism models is explicitly captured by some more complicated preference models. The reciprocity models, for instance, define an agent’s concern for others as varying according to her conjecture about their kindness to her (e.g., Rabin, 1993, Dufwenberg and Kirchsteiger, 2004, Falk and Fischbacher, 2006). These models usually involve the technical complexity of analyzing a psychological game. The inequity aversion models (e.g., Fehr and Schmidt, 1999, Bolton and Ockenfels, 2000), in contrast, employ a simpler idea to look at interactions by focusing on distributional (in)equality rather than intentions and beliefs. We extend this specific strand of literature in the current paper by taking an inequity aversion model to a dynamic game with simultaneous stage moves.

Experimental studies on public goods games have also increasingly been the subject during the past decades. Among various designs, the voluntary contribution mechanism that we look at in this study is examined extensively (see chapter 3 of Ledyard (1995) for a survey). It features several regularities, including: (i) contributions to the public goods in both one-shot and repeated games; (ii) levels of contribution start high but decline with repetitions; (iii) contributions plummet in the final rounds, but pick up in a surprise restart. Andreoni (1988, 1995) distinguishes altruism from confusion as the motivation
for individual contributions. He suggests that dissolved confusions are the reason of declining contributions. Palfrey and Prisbrey (1997), however, find the warm-glow effect may have a more significant influence than pure altruism. Croson (2007) notices a positive link between agents’ contributions as well as between one’s action and belief. She proposes conditional cooperation as an alternative hypothesis. This finding is then extended by several works. By first categorizing agents into conditional cooperators and selfish players, these works find cooperator groups always achieve high levels of public goods, whilst mixed groups with both types of agent would see declining cooperation along time (Gächter and Thöni, 2004, Burlando and Guala, 2005, Fischbacher and Gächter, 2008). Fischbacher et al. (2001) also notice that a majority of their conditional cooperative agents tend to contribute less than the group average. This self-biased conditional cooperation is then suggested as another cause for the patterns (Neugebauer et al., 2009). Finally, Fehr and Gächter (2000) (also Masclet et al. 2003, Sefton et al. 2007) show that if an agent could, she will punish free-riders in the group even at her own cost. It is suggested that this negative emotion associated with non-cooperative agents could be considered as a support for the reciprocity motivation.

The remaining parts of this paper are organized as following. Section 2 presents the model. In Section 3, the agents’ behavior in a static public goods game is studied. Section 4 presents the complete dynamic analyses, including both the complete information and incomplete information cases. Section 5 concludes. Proofs are all contained in the Appendix.

2 The Inequity Aversion Model

The inequity aversion model proposed by Fehr and Schmidt (1999) is a static game model that captures an individual’s concern beyond own material payoff in the following way:

$$u_i(x) = x_i - \alpha_i \frac{1}{n-1} \sum_{j \neq i} \max\{x_j - x_i, 0\} - \beta_i \frac{1}{n-1} \sum_{j \neq i} \max\{x_i - x_j, 0\}$$ (1)

In a group of $n$ agents, $i$ cares about not only her material payoff $x_i$, but also the payoff differences between herself and the others. This extra concern
is measured by $\alpha_i$ and $\beta_i$. Fehr and Schmidt suggest $0 \leq \beta_i < 1$ and $\alpha_i \geq \beta_i$, indicating that having either more or less payoff than others generates a psychological cost of inequity, whilst a disadvantageous inequity is more costly than an advantageous one.\footnote{This proposed value for $\beta$ rules out status seeking (one enjoys being better off than others, i.e., $\beta_i < 0$) or a willingness to sacrifice own payoff without improving the others’ ($\beta_i \geq 1$) in order to reduce an advantageous inequity.} The inequity among others is however not a direct concern.

For analytical simplicity, we will look at the $n = 2$ case in this paper. The two agents are labeled as agent 1 and 2. Since both the game and the preference model are linear in payoffs, we focus on uniform values of $\alpha$ and $\beta$ for inequity-averse agents without missing essential features of the model.\footnote{We will be dealing with threshold parameter values in our study. If parameters are of heterogeneous values, there will be anyway only two cases: either the target index composed of $\alpha_i$ and $\beta_i$ is beyond the threshold for an agent $i$ so that she behaves in the non-selfish way, or otherwise she acts not differently from a selfish one. Allowing the coexistence of selfish and inequity-averse agents equally captures these two cases.} A special attention will be paid to the case where inequity-averse and selfish agents coexist in a group, in order to find if and how they may cooperate with each other. To avoid confusions, we will refer to a selfish agent as “he” and denote him by $S$, and refer to an inequity-averse agent as “she” and denote her by $IA$.

The utility of an inequity-averse agent 1 can be expressed in the following form according to payoff differences:

$$\begin{align*}
    u_1 &= (1 + \alpha)x_1 - \alpha x_2, \quad \text{when } x_1 \leq x_2; \\
    u_1 &= (1 - \beta)x_1 + \beta x_2, \quad \text{when } x_1 > x_2.
\end{align*}$$

\[ (2) \]

\section{The Static Public Goods Game}

We start with the static public goods game. In this game, each agent receives an endowment $\omega$ and decides how much to contribute to the public goods. Denote their public goods contributions by $(g_1, g_2)$, with $g_i \in [0, \omega], i = 1, 2$. The rate of return to the public good is $a \in (1/2, 1)$, and that to the private
good is 1. An agent’s material payoff is \( x_i = \omega - g_i + a(g_1 + g_2) \), which gives an utility of

\[
u_1 = \omega + (a - 1)g_1 + ag_2 \tag{3}\]

if agent 1 is selfish, and

\[
u_1 = \begin{cases} 
\omega + (a - 1)g_1 + ag_2 - \alpha(g_1 - g_2), & \text{if } g_1 \geq g_2; \\
\omega + (a - 1)g_1 + ag_2 - \beta(g_2 - g_1), & \text{if } g_1 < g_2 
\end{cases} \tag{4}\]

if agent 1 is inequity-averse.

Notice that because the public goods is equally enjoyed by both agents, the inequity in payoffs reduces to the difference in contributions (we shorten “public goods contributions’ to “contributions” in the remaining of the paper). (3) indicates that for \( 1/2 < a < 1 \), contribution brings a pure cost to a selfish agent no matter the choice of the other agent. Therefore based on the selfish assumption, the traditional theory predicts \((0, 0)\) as the unique equilibrium, even though the material payoffs are maximized when both contribute \(\omega\). This is the typical free rider problem.

The first row in (4) demonstrates an even stronger disincentive for contribution if agent 1 is inequity-averse and expects to contribute more than the other. In this case, she will reduce \(g_1\) to save both the material and psychological cost. In the other case, when agent 1 expects to contribute less than the other, the second row in (4) shows that marginally increasing \(g_1\) incurs a material cost of \(1 - a\) but a psychological benefit of \(\beta\). For \(\beta\) large enough, the agent will optimally increase \(g_1\) until it matches \(g_2\). Such a behavior clearly distinguishes an inequity-averse agent from a selfish one. To keep our analyses interesting, we therefore assume through the paper that:

**Assumption 1.** \(a + \beta - 1 > 0\).

Fehr and Schmidt (1999) have a complete-information setup with \(k\) inequity-averse agents in a \(n\)-member group. We replicate their setup with our two-agent game by assuming that each agent only knows his/her own type but holds a common prior belief \(p\) as the possibility that the other is selfish.

An equilibrium is a pair of contributions \(g = (g_1, g_2)\) that defines the choice of either type, with which the utility of each \(agent\) is maximized. We
already know $g_s = 0$, and $g_{IA}$ is determined as following:

$$g_{IA} = \arg \max_{g_i} p[\omega + (a - 1)g_i - \alpha g_i] + (1-p)[\omega + (a - 1)g_i + a g_{IA} - \beta(g_{IA} - g_i)] \quad (5)$$

(5) implies that an inequity-averse agent will not deviate from an equilibrium by contributing less, for then she expects to end up contributing more than the other with probability $p$ and less than the other with probability $1 - p$. We do not need to consider upward deviations as we have discussed that an inequity-averse agent never contributes more than the other. The resulting equilibrium is then the following that resembles Fehr and Schmidt’s (1999) Proposition 4:

**Proposition 1.** It is always a dominant strategy for a selfish agent to contribute 0. For an inequity-averse agent, under the assumption that $a + \beta - 1 > 0$:

(a) If $1 - p < \frac{1 - a + \alpha}{\alpha + \beta}$, there is a unique equilibrium with $g_{IA} = g_s = 0$;

(b) If $1 - p \geq \frac{1 - a + \alpha}{\alpha + \beta}$, there is an infinity of equilibria. In each of these equilibria, $g_{IA} \in [0, \omega]$.

Proposition 1 is obtained through analyzing (5). It indicates that the model suffers from a severe multiple equilibria problem, as inequity-averse agents are willing to coordinate with each other on any level of contribution when the presence of selfish agents is sufficiently low. One can select certain equilibrium by using refinements such as payoff dominance or risk dominance à la Harsanyi and Selten (1988). Interested readers can find two of such exercises in the appendix.

We now take the inequity aversion model to the repeatedly played public goods game.

### 4 The Dynamic Public Goods Game

When the static game is repeated for $T$ times, it is conventionally called the “repeated public goods game” in the literature. Each repetition is referred to as a round. However, we shall notice that for inequity-averse agents, the payoff difference that they care about may not be an isolated variable across rounds, especially when we assume:
Assumption 2. *Inequity-averse agents only care about the difference in overall game payoffs.*

Under this assumption, an agent takes a global vision when choosing her strategy for the whole game. She may not worry much about the difference in contributions within a single round, as long as she can clear it by the end of game. We may also consider there being a stock variable called “contribution differential” that evolves across rounds. This finding allows us to claim that the repeatedly played game is not a repeated game, but rather a *dynamic* game. Hence we have the title of this paper.

This claim will sharply change the predictions. For example, if we have instead assumed that inequity-averse agents focus only on single round’s gains and losses, the game is indeed a repeated game. By employing backward induction, we then easily reach a prediction of zero contribution for all rounds if the group is mixed of both types plus that no type information is hidden.\(^6\) In contrast, we will see that assumption 2 leads to a much richer way to play the game in the equilibrium.

To characterize the dynamic game, we number rounds in the ascending order so that the game starts with round 1 and ends with round \(T\). Use set \(\{t\}\) to include all integers between 1 and \(T\), and set \(\{\omega\}\) for all real numbers between 0 and \(\omega\). \(A = A_1 \times A_2\) is the action set for the group of agents, with \(A_1 = A_2 = \{t\} \times \{\omega\}\). Each element of \(A\) is a vector \(g^t = (g_1^t, g_2^t)\), indicating a possible pair of public goods contributions by agent 1 and 2 at round \(t\). The complete realization of public goods contribution from round 1 to the start of round \(t\) makes the history of this round, which is denoted by \(h^t = (g^1, g^2, ..., g^{t-1})\). All the possible histories constitute the set of round \(t\)’s history \(H^t \in H\), where \(H = \{t\} \times H^t\). Let \(S_i\) be the set of pure strategy of agent \(i, i = 1, 2\). Each element in \(S_i\) is then a mapping from \(H\) to \(A_i\), denoted by \(s_i\). Finally, we may write down the intertemporal expected utility of an inequity-

\(^6\)Between this myopic assumption and our assumption 2, there are alternative assumptions that impose different bracketing sizes. For instance, we may assume that an agent has a perfect recall but no foresights, so that she tends to correct any contribution difference immediately. The dynamic game claim shall keep valid as long as the agents are not extremely myopic.
averse agent 1 at the beginning of any round $t$ as $u_1(s_1, s_2 | h^t) : H^t \times A \mapsto \mathbb{R}$:

\[
u_1^t(s_1, s_2 | h^t) = (t - 1) \cdot \omega - \sum_{r=1}^{t-1} g_1^r + a \left( \sum_{r=1}^{t-1} g_1^r + \sum_{r=1}^{t-1} g_2^r \right) +
\]

\[
+ (T - t + 1) \cdot \omega - \sum_{r=t}^{T} g_1^r + a \left( \sum_{r=t}^{T} g_1^r + \sum_{r=t}^{T} g_2^r \right) -
\]

\[
- \alpha \max \left\{ \sum_{r=1}^{T} \bar{g}_1^r - \sum_{r=1}^{T} \bar{g}_2^r, 0 \right\} - \beta \max \left\{ \sum_{r=1}^{T} \bar{g}_2^r - \sum_{r=1}^{T} \bar{g}_1^r, 0 \right\}
\]

The first row in (6) describes agent 1’s acquired material payoff until the start of round $t$; the second row is her anticipated material payoff from round $t$ on until the end of game, according to her strategy and that of the other agent. Upper-bars are for anticipated values. The third row corresponds to our assumption and records the anticipated aggregate payoff difference from the entire game. Notice that the psychological cost of inequity arises if and only if the two agents finish the game with different aggregate contributions to the public goods.

4.1 The Game with Complete Information

Before examining the complete scenario where different types are mixed in a group under hidden type information, we first look at the complete information scenarios. They will allow us to have a clean view of the impact of inequity aversion by isolating any reputation effect.

By employing backward induction, we have a prediction for the case where both agents are selfish: since neither agent contributes in the final round whatever the history, round $T$ cannot be used as a credible threat to maintain cooperation in the penultimate round, which has to end up with $g^{T-1} = (0, 0)$. Rolling this backwards to the first round gives a unique equilibrium of zero contribution, i.e., $(g_1^t, g_2^t) = (0, 0), \forall t$.

\footnote{Obviously, (6) is obtained by replacing instantaneous material payoffs $x_1^t$ and $x_2^t$ in:

\[
u_1^t = \sum_{r=1}^{t-1} x_1^r + \sum_{r=t}^{T} \bar{x}_1^r - \alpha \max \left\{ \sum_{r=1}^{T} \bar{x}_2^r - \sum_{r=1}^{T} \bar{x}_1^r, 0 \right\} - \beta \max \left\{ \sum_{r=1}^{T} \bar{x}_1^r - \sum_{r=1}^{T} \bar{x}_2^r, 0 \right\}.
\]
The other case of homogeneous group is when both agents are inequity-averse. Based on assumption 2, we may treat the whole game (from round 1 to \( T \)) as a static game with \( p = 0 \). Part (b) in Proposition 1 tells us that given \( a + \beta - 1 > 0 \), the fear for incurring any psychological cost of inequity shall drive agents to perfectly coordinate their aggregate contributions throughout the game.

The question is then about the coordinated level of contributions. Proposition 2 shows that in the equilibrium, this is defined in the form of minimum cooperation in any remaining game:

**Proposition 2.** For the dynamic public goods game played between two inequity-averse agents, given \( a + \beta - 1 > 0 \), a pure strategy subgame perfect equilibrium, \((g^1_t, g^2_t)\), is fully characterized by:

(a) \( \sum_{i=1}^{T} g^i_1 = \sum_{i=1}^{T} g^i_2 \),

(b) \( \sum_{r=t}^{T} g^r_i \geq (T - t)\omega, \forall t \in [1, T], i = 1, 2 \).

The proofs for propositions including this one are relegated to the appendix and we convey the intuitions here. Part (a) shows the agents’ perfect coordination on aggregate contributions, due to assumptions 1 and 2. The whole game is like a single round here, where only aggregate contributions (equivalently, aggregate payoffs) matter for maintaining equity.

Part (b) shows that there is a minimum level of cooperation in the remaining game of any round. The intuition behind is simple, that is an important enough future is necessary to regulate behavior in any round of the game. Because in this way, no one will want to make a deviation and forfeit the future. It turns out that this minimum level is below the efficient contributions by only \( \omega \).

We can use the following two-round game to demonstrate the mechanism.

Suppose \( g^1 = (g^1_1, g^1_2) \) and \( g^2 = (g^2_1, g^2_2) \) are agents’ contributions in an equilibrium, with \( g^1_1 + g^2_1 = g^1_2 + g^2_2 \). If agent 1 deviates to contributing \( \omega \) rather than \( g^1_1 \) in the first round, the worst case that she can have for the second round is \( \hat{g}^2 = (0, \omega - g^1_2) \): in this way the aggregate contributions are

---

\(^8\)Because at any round \( t \), the efficient level of contribution aggregated from this round to the final round is \( (T - t + 1)\omega \).
maintained equal, meanwhile the public goods is realized on the lowest possible level. The deviation is hence not profitable if

\[ [\omega - \omega + a(\omega + g_2)] + [\omega + a(\omega - g_2)] \leq [\omega - g_1 + a(g_1 + g_2)] + [\omega - g_1 + a(g_1 + g_2)] \] (7)

which by applying \( g_1 + g_2 = g_2 + g_2 \) reduces to

\[ g_i^1 + g_i^2 \geq \omega, \quad i = 1, 2. \] (8)

This is nothing but the \( T = 2 \) version of part (b) in Proposition 2 when \( t = 1 \).

This mechanism is guaranteed to be valid as agents are committed to match each other’s action at any aggregate level. For \( T > 2 \), one just needs to pay attention that the deviation punishments (through forfeiting future public goods) must ensure subgame perfection for any remaining game.

Notice that when \( t = 1 \), part (b) determines the minimum public goods contributions in the overall game. An agent’s total contribution to the public goods is then at least \((T - 1)\omega\). It is a surprisingly efficient result, especially recalling that any level of symmetric contribution from 0 to \( \omega \) can appear in an equilibrium of a static game. The dynamic game apparently functions as a very efficient coordination device for the inequity-averse agents. What is even more significant is that, as we will shortly demonstrate, this coordination device is equally effective when selfish agents join the group.

Proposition 2 does not tell what agents do in each individual round, although it does imply a reduced multiple equilibria problem from the case should the game be played as a repeated one. In an attempt to further reduce the size of the equilibrium set, I choose to curb the severity of punishments for deviations. This will allow more deviations to become profitable.

For this purpose, we first define the punishment. In order to highlight agents’ intention to match each other’s action, we can define punishment as the (induced) contribution reduction of the deviating agent in the remaining game, after her deviation has been (feasibly) matched based on the equilibrium

---

\(^9\)To see this, just set \( p = 0 \) in part (b) of Proposition 1.

\(^{10}\)If it was a repeated game, what happens in a static game may equally happen in each round, so that we expect a severe multiple equilibria problem with symmetry in each round at any level of contribution.
path. To illustrate this concept, denote the equilibrium contributions by \( g^t = (g^t_1, g^t_2), t \in [1, T] \). Without loss of generality, let player 1 deviate to \( g^k_1 + \Delta \) in a round \( k \), where \( \Delta \in [-g^k_1, \omega - g^k_1] \). Following this deviation, the remaining game is then played with contribution pairs \( \tilde{g}^r = (\tilde{g}^r_1, \tilde{g}^r_2), \forall r > k \), with which:

**Definition 1.** The punishment \( P \) is:

- \( P = \sum_{r=k+1}^{T} g^r_1 - \sum_{r=k+1}^{T} \tilde{g}^r_1, \) if \( \sum_{r=k+1}^{T} g^r_2 + \Delta \in [(T-k-1)\omega, (T-k)\omega] \);

- \( P = \sum_{r=k+1}^{T} g^r_1 - \left[ \sum_{r=k+1}^{T} g^r_2 + \Delta - (T-k)\omega \right] - \sum_{r=k+1}^{T} \tilde{g}^r_1, \) if \( \sum_{r=k+1}^{T} g^r_2 + \Delta \in ((T-k)\omega, \sum_{r=k+1}^{T} \tilde{g}^r_1 + \omega] \);

- \( P = 0, \) if \( \sum_{r=k+1}^{T} g^r_2 + \Delta > \sum_{r=k+1}^{T} g^r_1 + \omega \) or \( \sum_{r=k+1}^{T} g^r_2 + \Delta < (T-k-1)\omega \)

The last two cases correspond to the feasibility argument. For now if agent 2 fully matches the deviation before imposing any punishment, her aggregate contribution would have gone beyond the feasible range (i.e., \([ (T-k-1)\omega, (T-k)\omega] \) for the remaining game of round \( k \), taking into account the requirement of subgame perfection). The punishment is then imposed only after the agents have coordinated to adjust for the infeasible part.\(^{11}\) In this way, 0 punishment is always well defined.

The harshest punishment is of course the one that keeps the aggregate contribution in the remaining game at the subgame perfect minimum (i.e., \( \sum_{r=k+1}^{T} \tilde{g}^r_1 = (T-k-1)\omega \)). This is indeed what we have used to obtain the largest set of subgame perfect equilibrium in Proposition 2. Now I introduce a Limited Punishment for Deviations (LPD) principle, which supposes agents agree to control the size of punishment to be smaller than the deviation:

**Limited Punishment for Deviations (LPD).** The punishment for any deviation is held smaller than the deviation itself, i.e., \( P < |\Delta| \).

\(^{11}\)Let us take the second case as an example. Now \( \sum_{r=k+1}^{T} g^r_2 + \Delta > (T-k)\omega \), which makes it impossible for agent 2 to directly increase her aggregate contribution starting from round \( k+1 \) by \( \Delta \). Imagine instead the following decomposed moves: agent 2 first adjusts the desired contribution from \( \sum_{r=k+1}^{T} g^r_2 + \Delta \) to \( (T-k)\omega \). Agent 1 then matches this adjustment to keep aggregate contributions equal. Agent 1’s aggregate contribution in the remaining game is now \( \sum_{r=k+1}^{T} g^r_1 - [\sum_{r=k+1}^{T} g^r_2 + \Delta - (T-k)\omega] \). Since it is still higher than the subgame perfect minimum (= \( (T-k-1)\omega \)), a positive punishment can then be imposed.
\(|\Delta|\) indicates the deviation’s absolute value, and \(P\) is as defined in Definition 1. This punishment ceiling makes deviations more likely to be profitable and therefore downsizes the equilibrium set. Specifically, the equilibria under LPD are:

**Proposition 3.** Under LPD, a pure strategy equilibrium of the dynamic game played between two inequity-averse agents is featured with the following contributions:

(a) \(g^t_1 = g^t_2, \forall t \in [1, T]\);

(b) \(g^t_1 = g^t_2 = \omega, \forall t \in [1, T]\) with at most one exceptional \(k \in [1, T]\), when \(g^k_1 = g^k_2 \in [0, \omega]\).

Proposition 3 predicts that agents coordinate contributions in each individual round; and with at most one exceptional round, they always contribute at the efficient level. The LPD principle hence further enhances the coordination function of the dynamic game. The single round contribution levels are now such because whenever there are more than one round with contributions less than maximum, deviating to contribute more in the earlier round is profitable under the protection of LPD. At the same time, imposing such a restriction on the punishment with LPD is however not as arbitrary as it seems to be. On the one hand, punishment is a sword with double blades: it reduces not only the public goods enjoyment of the deviator, but also that of the punisher. On the other hand, letting one suffer (not more than) what he has caused seems to also accord with common social practices. Beside these reasons, LPD is apparently sufficient to realize the purpose of alleviating the multiple equilibria problem.

Having seen that inequity-averse agents always achieve highly efficient cooperation with each other, we wonder what happens if the group consists of an inequity-averse agent and a selfish one. Will we be able to use backward induction to draw a conclusion of zero contributions, because the selfish agent does not cooperate in the final round? The answer is No, again thanks to the coordination function of the dynamic game. Even knowing that the selfish agent does not contribute in the final round, the inequity-averse agent may
still contribute in this round as long as it reduces the payoff inequity. This final-round contribution can then be used as a threat to maintain contributions in previous round(s). And the threat itself is credible so long as the selfish agent has made a larger aggregate contribution before the final round.

A more concrete description of the equilibrium contributions is made in the following proposition:

**Proposition 4.** With public information about agents’ types, a mixed group’s contributions in the dynamic game follow \( \sum_{t=1}^{T} g_{IA}^t = \sum_{t=1}^{T} g_{S}^t, \quad g_{S}^T = 0, \) and

\[
\text{Without LPD:} \quad \sum_{t=1}^{T-1} g_i^t \geq (T - 1 - t)\omega, \quad i = IA, S, \forall t \in [1, T - 1], \text{ and:}
\]

- either \( g_{IA}^T = \omega, \)
- or \( g_{IA}^T \in [1 - \frac{1}{a}\omega, \omega], g_{S}^{T-1} = \omega. \)

\[\text{With LPD:} \quad g_{S}^t = \omega, \forall t < T; g_{IA}^T \in [1 - \frac{1}{a}\omega, \omega].\]

Proposition 4 highlights again the perfect match in aggregate contributions of the two agents. It also clearly shows the importance of a large enough final-round contribution for the threat to be effective. As the inequity-averse agent provides an external incentive for the selfish agent to contribute, it is apparently also to the latter’s interest to exploit this mechanism. Notice that the selfish agent even contributes more aggressively than the inequity-averse agent till the final round. Such an “opportunistic generosity” is guaranteed to be payed back by the inequity-averse agent’s commitment to match contributions. Nevertheless, because the selfish agent does not contribute at the end of the game, the mixed group never reaches a higher level of efficiency than a homogeneous group of inequity-averse agents. This can be clearly seen when we compare the levels of contribution in Proposition 4 to those in Proposition 3.

This possibility to form nearly efficient cooperation between selfish and non-selfish agents under complete information clearly does not exist for a finite repeated game. It actually helps to explain the restart effect recorded in public goods experiments. The restart effect refers to the observation that contributions recover to high levels after having dropped at the end of the
game, if agents are unexpectedly asked to play the game again. As Proposition 3 and 4 imply, this may be caused by the cooperation between agents whose type information has been completely revealed.

4.2 The Game with Incomplete Information

In the end, we complete this study by looking at the situation where the type information is private. Each agent holds the common prior belief \( p \) that the other is selfish. This belief is updated as the game is being played. We denote the updated belief at the beginning of round \( t \) by \( p(h^t) \), or \( p^t \) for notational simplicity, with \( p^1 = p \). The type \( i \) agent’s \((i = IA, S)\) intertemporal expected utility at round \( t \) is \( u^t_i(s_{IA}, s_S|h^t, p^t) \) if agents play pure strategies or \( u^t_i(\sigma_{IA}, \sigma_S|h^t, p^t) \) if agents play mixed strategies, where \( \sigma_i \) denotes \( i \)'s mixed strategy. We suppose that the LPD principle is always adopted.

When agents play pure strategies in the game, it may happen that an agent never contributes anything if expecting to play against a selfish one at a high probability (i.e., \( p \) is large). Because then if one tries to deviate from a zero contribution equilibrium, the expectation that the deviating contribution be matched is small. This means the deviation is not likely to be profitable, especially for an inequity-averse agent who risks an extra psychological cost of disadvantageous inequity.

However, a small \( p \) could induce both types of agent to contribute equally and maximally until very end of the game. Because then even without any update of belief, an inequity-averse agent is still willing to make contribution in the final round, as is shown in the part (b) of Proposition 1. A contribution equilibrium would only require \( g_{IA}^T \) to be large enough so that it can be used as a threat to support early rounds’ contributions.

It may also be the case that beliefs are updated before the game ends. Suppose the two types of agent distinguish from each other by making different levels of contribution starting from a specific round, e.g., \( g_{IA}^k > 0 \) and \( g_S^k = 0 \) for \( k \in [1, T] \). They then play the remaining game under complete information. It turns to be however difficult for such a separation to happen in an early stage, because then a revealed selfish agent needs to catch up in contribution before he may benefit again from cooperation. A late separation of types is
hence more favorable with a longer period of full cooperation. We formalize these discussions in the following Proposition.

**Proposition 5.** For a game of $T$ rounds played under incomplete type information, contributions in a pure strategy equilibrium are realized as following. (a) and (b) are also unique under LPD:

(a) for $1 - p < \min \left\{ \frac{1-a}{a}, \frac{1-a+a}{a+\beta} \right\}$: $g^t = (0,0), \forall t \in [1, T]$;

(b) for $1 - p \geq \max \left\{ \frac{1-a+a}{a+\beta}, \frac{1-a+a}{a+\alpha} \right\}$:

- either $g^t = (\omega, \omega), \forall t < T$, and $g^T_{IA} \in \left[ \frac{1-a}{a}, \frac{\omega}{1-p} \right], g^T_S = 0$;

- or $g^k_S = g^T_{IA} = \omega, \forall t < T$ and $t \neq k$; $g^k_S = g^k_{IA} \in [0, \omega]$; $g^T_S = 0, g^T_{IA} = \omega$.

(c) for $1 - p \geq \frac{(1-a)^2+aa}{a(a-1)+a} : g^t = (\omega, \omega), \forall t < T - 1$. $g^{T-1}_{IA} = \omega, g^{T-1}_S = g^T_S = 0$.

And $g^T_{IA} \in \left[ \frac{(1-a+a)p}{(2a-1)(1-p)}, \frac{(1-a) \omega}{a+\omega} \right]$ if $g^{T-1} = (\omega, \omega)$, otherwise $g^T_{IA} = 0$.

(a) and (b) contain all the pooling equilibria and (c) contains the separating equilibria. The value range of $g^T_{IA}$ chosen for a separating equilibrium reflects the necessity to prevent deviations by either type of agent. On the one hand it is bounded from above, so that a selfish agent does not find enough benefit to mimic the inequity-averse type. On the other hand it is bounded from below, so that an inequity-averse agent feels facing a significant loss if missing it by not revealing herself. Interestingly, we realize that an inequity-averse agent risks a disadvantageous inequity by contributing in the penultimate round to distinguish from a selfish agent. She may alternatively avoid all psychological costs if she waits for the other to contribute in the separating round, then catches up in the final round. This potential “wait-and-see” deviation indeed puts an even higher requirement on the lower bound of $g^T_{IA}$. As a result, the final round contribution must be larger than 0, but smaller than $\omega$.\(^{12}\)

The final round’s contributions, as defined in (b) and (c) in above, help to explain the end-game effect observed in lab experiments. This effect refers to a sharp drop of contributions when the game approaches the end. By assuming

\(^{12}\)To prevent an inequity-averse agent from playing the wait-and-see deviation, the devi-
agents play a repeated game, Fehr and Schmidt (1999) predict that contributions are zero in the final round, because their calibrated parameter values cannot support contributions in a static game. However, as the experimental data frequently show that contributions stay low but still positive in the final round, our dynamic game “prediction” may be closer to the reality.

When agents play mixed strategies in the game, the zero contribution prediction when \( p \) is large can be revised. As Kreps and Wilson (1982) show in their seminal paper on the chain-store paradox, even when there is little suspect about an incumbent monopolist being irrational and always fighting entry, market entries will happen only at a late stage of a finitely repeated game. This is because a rational monopolist will mimic the irrational’s behavior by also fighting entries. Kreps et al. (1982) use the same reputation game mechanism to examine the cooperation in a finitely repeated prisoners’ dilemma. They show that the suspected existence of a “Tit-for-Tat” player is sufficient for the cooperation to form.

Agents in the current game may just exploit the same mechanism. For example, imagine that inequity-averse agents stick to contributing \( \omega \) meanwhile

\[
D_{tA} = \left[ \omega + a(1-p)g_{tA}^{T-1} + \omega + (a-1)(1-p)g_{tA}^{T-1} \right] - \left[ \omega + (a-1)g_{tA}^{T-1} + a(1-p)g_{tA}^{T-1} + \omega + (2a-1)(1-p)g_{tA}^{T-1} - \alpha pg_{tA}^{T-1} \right] \\
= (1-a+\alpha)pg_{tA}^{T-1} - (2a-1)(1-p)g_{tA}^{T-1} \leq 0. \tag{9}
\]

In the last row, the first item indicates the material cost and the psychological cost that are saved if she chooses to wait. This saving is realized if the other is revealed as selfish, which happens with a probability of \( p \). The second item is the forfeited cooperation in round \( T \) that is relevant when the other is revealed as inequity-averse, which has a probability of \( 1-p \). (9) clearly requires a large enough \( g_{tA}^{T} \) with regard to \( g_{tA}^{T-1} \). For a selfish agent, the no deviation condition is

\[
D_s = \left[ \omega + (a-1)g_{tA}^{T-1} + a(1-p)g_{tA}^{T-1} + \omega + a(1-p)g_{tA}^{T} \right] - \left[ \omega + a(1-p)g_{tA}^{T-1} + \omega \right] \\
= (a-1)g_{tA}^{T-1} + a(1-p)g_{tA}^{T} \leq 0 \tag{10}
\]

Hence it requires \( g_{tA}^{T} \) to be small enough with regard to \( g_{tA}^{T-1} \). Combining (9) and (10) shows that \( g_{tA}^{T} = \omega \) is never acceptable in the equilibrium.
selfish agents randomize between contributing $\omega$ and 0. Observing an agent being contributing $\omega$ all the time will gradually convince the other that this is indeed an inequity-averse agent. Contributions may then be maintained even in the final rounds, regardless of the low prior belief about the existence of inequity-averse agents.

In fact, the reputation game will be played somewhat in the opposite way under the inequity aversion model. Let us focus on $\omega$ and 0 as the contribution choices in the game. Denote $r^t$ as the probability that a selfish agent contributes $\omega$ in round $t$, and $q^t$ as the probability that $\omega$ is contributed by an agent in the same round. So if an inequity-averse agent is sure to contribute $\omega$ in this round, there is $q^t = (1 - p^t) + p^t r^t$. The belief shall be updated along the game according to the Bayes’ rule, namely $1 - p^{t+1} = (1 - p^t)/q^t$. Whenever the selfish agents do not contribute $\omega$ for certain, i.e., $r^t < 1$, one will assign a larger $1 - p^{t+1}$ than $1 - p^t$ to his/her partner after observing $\omega$ in round $t$.

Here is a way to play the game. When it starts, both agents contribute $\omega$ till round $k - 1$ inclusive, where $k = T - (2n - 1)$ for some $n \in \mathbb{N}$. In round $k$, a selfish agent randomizes between contributing $\omega$ and 0, meanwhile an inequity-averse agent contributes 0. When $g^k = (\omega, 0)$, the one who contributes $\omega$ is revealed as selfish and the one who contributes 0 is believed more likely to be inequity-averse than before. We denote the revealed selfish agent by $S_R$ and an unrevealed selfish agent by $S_{UR}$.

If $g^k \neq (\omega, 0)$, then both agents contribute 0 forever. If $g^k = (\omega, 0)$, then in each round that follows, an inequity-averse agent matches $S_R$’s contribution of the previous round. An $S_{UR}$ agent mimics the inequity-averse agent by randomizing between contributing $\omega$ and 0, as long as $S_R$ has always contributed $\omega$. In round $T - 1$ and $T$, $S_{UR}$ contributes 0. As for $S_R$, he first randomizes between contributing $\omega$ and 0 in round $k + 1$. Then he alternatively plays contribution and randomization as long as $(\omega, \omega)$ has always been contributed by the group. In round $T$, $S_R$ also contributes 0. However, whenever the group fails to contribute $(\omega, \omega)$ in a round, both $S_{UR}$ and $S_R$ will contribute 0 forever. We claim that

**Proposition 6.** The way to play the game as described above follows a subgame perfect equilibrium.
We briefly discuss several noteworthy features of this equilibrium here, and leave its elaboration to the appendix. First notice that when the belief starts to be updated, it is the selfish agent who maintains the contribution. The inequity-averse agent in the meantime plays “wait-and-see”, because the low prior belief about $1 - p$ makes the risk of disadvantageous inequity too high for her to distinguish herself. Instead, waiting for the other to contribute and then catching up ex post allows to avoid any unequal result. Secondly, only a selfish agent who is not revealed at the first place has the opportunity to build a reputation by mimicking the inequity-averse type. Meanwhile, the leading contributor $S_R$ has to make sure that the other agent’s catch-up in contribution is not always rewarded with future cooperation. Because only in this way would an $S_{UR}$ agent have the incentive to gradually reveal himself in the remaining game.

This equilibrium shows one possibility that the public goods are contributed efficiently in the early stage of a game. The prior belief matters for how long this stage can be. Basically, the smaller is $1 - p$, the longer would be the second stage for the belief to be sufficiently updated, so that desired contributions can be supported in the end of game. At the same time, the “wait-and-see” strategy of an inequity-averse agent also suggests fluctuations in contribution that are commonly observed in the experiments.

5 Conclusions

In this paper we have examined the inequity aversion model in the public goods game. The model highlights individuals’ aversion to distributional inequity that accompanies their normal material enjoyment. An inequity-averse agent hence has an incentive to match others’ behavior in order to achieve equal payoffs. Their behavior exhibits a conditional cooperation pattern in the public goods game.

In order to apply the originally static model to a finitely repeated version of the public goods game, we assume that agents focus on the inequity between anticipated overall payoffs. This assumption excludes any form of myopia, and renders the current game to be treated as a dynamic game instead of a repeated
game. We hence challenge the common perspectives in the literature, and obtain results with highly efficient public goods contributions in each subgame perfect equilibrium. However, the predictions for individual round’s contributions are obtained only after a behavior rule called Limited Punishment for Deviations (LPD) is introduced. Under LPD, we predict nearly maximum cooperation between inequity-averse agents.

In the dynamic game, such cooperation is highly maintained even when selfish agents join the group. Committed to match overall contributions even in the final round, an inequity-averse agent provides an incentive for a selfish agent to cooperate until the game approaches the end. This feature keeps qualitatively unchanged when informational asymmetry is introduced over agents’ types. In that case, a selfish agent plays the role of a more active cooperator.

Beyond cooperation, we also manage to capture some renowned regularities of the public goods game. These especially include the end-game effect and the restart effect. The former closely relates to the revelation of agents’ types in a game with asymmetric information, and the latter relies on the formation of cooperation under complete information, provided that at least one agent of the group is averse to inequity.

We shall admit that the current model does not encompass the declining contributions in the game. With regard to this trend a recent literature has proposed several explanations. We close this paper by quoting two of these studies for interested readers. Ambrus and Pathak (2009) suggest to introduce “reciprocal” agents whose only concern is to match actions to a group average. This group average is obtained from the beginning of game till the current round. If the group consists of both selfish and reciprocal agents, the average contribution will start to drop when the selfish ones withdraw their contributions as less rounds remain. The reciprocal agents then tune down their contributions accordingly, so that the declining pattern appears. Figuières et al. (2009) introduce “weakly morally motivated” agents who tend to comply behavior with their moral objects that are influenced by others’ behavior. This object is updated in each round based on one’s prior ethical level and the previous round’s group performance. The tension between material costs of contribution and psychological costs of not attaining the object determines
that one always contributes below her object. This then leads to a general decay of the moral objects in the group as the game is played, which in turn causes the contributions to also decline over time.

References


Appendices

A Equilibrium Selection in a Static Game

In this section we demonstrate two equilibrium refinements for the static game played by a mixed group under hidden type information. Namely Payoff Dominance (PD), and Risk Dominance (RD). These refinements are proposed by Harsanyi and Selten (1988) to guide the coordination on one of many equilibria.

Payoff Dominance. A payoff dominant equilibrium is one that is Pareto superior to any other equilibria in the game. Based on Proposition 1, for $a + \beta - 1 > 0$, an inequity-averse agent’s payoff in an equilibrium is:

$$u_{IA}(g_{IA}, 0) = p[\omega - (1 - a)g_{IA} - \alpha g_{IA}] + (1 - p)[\omega - (1 - a)g_{IA} + ag_{IA}]$$

$$= \omega - [(2a - 1) - (a + \alpha)p]g_{IA}$$

The payoff dominant equilibrium is hence $g_S = 0$, and:

$$g_{IA} = \begin{cases} 
\omega, & \text{if } p < \frac{2a-1}{a+\alpha} \\
0, & \text{otherwise}
\end{cases}$$

Note the condition for positive contributions is not the same as part (b) in Proposition 1.

Risk Dominance. We refer to a risk dominant equilibrium as one that gives the highest expected utility when agent 1 thinks that an inequity-averse agent 2 would choose any equilibrium level $g_2$ according to a uniform distribution between 0 and $\omega$, and vice versa for agent 2. An inequity-averse agent 1 then chooses $g_1$ to maximize:

$$u_1 = p[\omega - (1 - a)g_1 - \alpha g_1] +$$
$$+ (1 - p)\{Prob(g_1 \geq g_2)[\omega - (1 - a)g_1 + ag_2 - \alpha(g_1 - g_2)] +$$
$$+ Prob(g_1 < g_2)[\omega - (1 - a)g_1 + ag_2 - \beta(g_2 - g_1)]\}$$

Treat the distribution as if it is continuous and re-present the objective function
as

\[ u_1 = p[\omega - (1 - a)g_1 - \alpha g_1] + \]
\[ + (1 - p)\left\{ \int_0^{g_1} \frac{1}{\omega} [\omega - (1 - a)g_1 + ag_2 - \alpha(g_1 - g_2)]dg_2 + \right\} \]
\[ + \int_0^\omega \frac{1}{\omega} [\omega - (1 - a)g_1 + ag_2 - \beta(g_2 - g_1)]dg_2 \],

which reduces to:

\[ u_1 = \frac{1 - p}{\omega} (\alpha + \beta)g_1^2 + [a - 1 - \alpha + (1 - p)(\alpha + \beta)]g_1 + \omega + \frac{1}{2}(1 - p)(a - \beta)\omega \]

The first order condition then gives the RD equilibrium contribution with

\[ g_S = 0 \] and

\[ \begin{cases} g_{1A} = \frac{(1-p)(\alpha + \beta) - (1-a+\alpha)}{(1-p)(\alpha + \beta)} \omega, & \text{if } p < \frac{a+\beta-1}{\alpha+\beta}, \\ g_{1A} = 0, & \text{otherwise.} \end{cases} \]

\section*{B Prove Propositions}

\subsection*{B.1 Prove Proposition 2 and Proposition 3}

Proposition 2 and 3 describe equilibrium contributions when homogeneous inequity-averse agents play the dynamic game under complete information, with and without the LPD principle. We prove them in below.

\textbf{Lemma 1.} \textit{There is no pure strategy equilibrium where }\sum_{t=1}^{T} g_1^t \neq \sum_{t=1}^{T} g_2^t.\textit{ }

\textit{Proof.} Use contradictions. Suppose } g^t \text{ is an equilibrium with } \sum_{t=1}^{T} g_1^t - \sum_{t=1}^{T} g_2^t = \epsilon > 0, \text{ without loss of generality. We start by looking at the final round.}

(I) \textit{Suppose } g_1^T > 0. \textit{Applying (6) to } t = T, \textit{agent 1 if deviates to } \hat{g}_1^T = \max\{g_1^T - \epsilon, 0\} \text{ obtains a net benefit of deviation (denoted by } D_1):}

\[ D_1 = [\omega + (a - 1)(g_1^T - \epsilon) + ag_2^T] - [\omega + (a - 1)g_1^T + ag_2^T - \alpha\epsilon] \]
\[ = (1 - a + \alpha)\epsilon, \text{ if } g_1^T - \epsilon > 0, \text{ otherwise:} \]

\[ D_1 = [\omega + ag_2^T - \alpha(\epsilon - g_1^T)] - [\omega + (a - 1)g_1^T + ag_2^T - \alpha\epsilon] \]
\[ = (1 - a + \alpha)g_1^T. \]
We have cancelled the payoffs received before round-$T$ in above as they are not affected by the deviation.

Therefore to avoid such deviations, an equilibrium must have either $\sum_{t=1}^{T} g_1^t = \sum_{t=1}^{T} g_2^t$ (so $\epsilon = 0$), or $\sum_{t=1}^{T} g_1^t - \sum_{t=1}^{T} g_2^t = \epsilon > 0$ but $g_1^T = 0$. Let us further suppose $g_2^T < \omega$. Then agent 2 if deviates to $g_2^T = \min\{g_2^T + \epsilon, \omega\}$, will have a net benefit (thanks to assumption 1) of deviation:

$$D_2 = [\omega + (a - 1)(g_2^T + \epsilon) + ag_1^T] - [\omega + (a - 1)g_2^T + ag_1^T - \beta \epsilon]$$

$$= (a + \beta - 1)\epsilon, \text{ if } g_1^T + \epsilon < \omega, \text{ otherwise:}$$

$$D_2 = [\omega + (a - 1)\omega + ag_1^T - \beta (\epsilon - (\omega - g_2^T)))] - [\omega + (a - 1)g_2^T + ag_1^T - \beta \epsilon]$$

$$= (a + \beta - 1)(\omega - g_2^T).$$

Combining the findings for both agent 1 and 2, we know if an equilibrium has $\sum_{t=1}^{T} g_1^t - \sum_{t=1}^{T} g_2^t = \epsilon > 0$, it must also have $g_1^T = 0, g_2^T = \omega$.

(II) The above analyses equally apply to round $T - 1$. Just notice that with $g_1^T = 0, g_2^T = \omega$, there is no room for punishment (see Definition 1) following an agent’s deviation in round $T - 1$. Therefore in an equilibrium where $\sum_{t=1}^{T} g_1^t - \sum_{t=1}^{T} g_2^t = \epsilon > 0$, there must be $g_1^{T-1} = g_1^T = 0, g_2^{T-1} = g_2^T = \omega$.

(III) Use backward induction to the beginning of game, a contradiction between $\sum_{t=1}^{T} g_1^t - \sum_{t=1}^{T} g_2^t = \epsilon > 0$ and $g_1^t = 0, g_2^t = \omega, \forall t$, is reached. The Lemma is proved. $\square$

**Lemma 2.** In a $T$ round game ($T \geq 2$), all equilibrium feature $\sum_{r=t}^{T} g_i^r \geq (T-t)\omega, \forall t \in [1, T], i = 1, 2$.

**Proof.** Again use contradictions. We show that if in an equilibrium $\exists t \in [1, T]$ with $\sum_{r=t}^{T} g_i^r < (T-t)\omega$, deviations will be made. Define the equilibrium strategy $g_i^r, i = 1, 2$:

(i) $\sum_{t=1}^{T} g_1^t = \sum_{t=1}^{T} g_2^t$;

(ii) If $g_k^r = (g_1^r + \Delta, g_2^r), \forall k \in [1, T-1], \forall \Delta \in [-g_1^k, \omega - g_1^k]$, then $g^r = (\hat{g}_1^r, \hat{g}_2^r), \forall r \in [k + 1, T]$, which satisfies:

(ii.a) $\sum_{r=k}^{T} \hat{g}_1^r = (T-t)\omega$ and $\sum_{r=k}^{T} \hat{g}_2^r \geq (T-t)\omega, \forall t \in [k + 1, T]$;

$\sum_{r=k+1}^{T} \hat{g}_2^r = (T-k-1)\omega + G^{k+1}$, if $G^{k+1} \in [0, \omega]$;

30
(ii.b) \( \sum_{r=t}^{T} \hat{g}_1^i \geq (T-t)\omega \) and \( \sum_{r=t}^{T} \hat{g}_2^i = (T-t)\omega, \forall t \in [k+1, T]; \) 
\( \sum_{r=k+1}^{T} \hat{g}_1^i = (T-k-1)\omega - G^{k+1} \) if \( G^{k+1} \in [-\omega, 0); \) 
(ii.c) \( \sum_{r=t}^{T} \hat{g}_1^i = (T-t)\omega \) and \( \sum_{r=t}^{T} \hat{g}_2^i = (T-t+1)\omega, \forall t \in [k+1, T], \) if \( G^{k+1} > \omega; \)
(ii.d) \( \sum_{r=t}^{T} \hat{g}_1^i = (T-t+1)\omega, \sum_{r=t}^{T} \hat{g}_2^i = (T-t)\omega, \forall t \in [k+1, T], \) if \( G^{k+1} < -\omega. \)

\( G^{k+1} = \sum_{r=k+1}^{T} g_2^i - \sum_{r=k+1}^{T} g_1^i + \Delta. \) Part (ii) includes all scenarios that make the harshest punishment for a deviation in an equilibrium (by Definition 1 in the text, e.g., \( P = \sum_{r=k+1}^{T} g_1^i - (T-k-1)\omega \) for (ii.a) and \( P = \sum_{r=k+1}^{T} g_2^i + \Delta - (T-k-1)\omega \) for (ii.b)). They are the harshest because following a deviation, the public goods realization is minimized with at least one agent’s contribution being driven to the minimal subgame perfect level (i.e., \( (T-k-1)\omega \)). At the mean time, they keep aggregate contributions as equal as possible between players.\(^{13}\)

The proof starts with supposing \( g_1^{T-1} + g_1^T < \omega, \) i.e., \( t = T - 1. \) Let agent 1 take the maximal \( \Delta > 0 \) deviation in round \( T - 1, \) i.e., \( \Delta = \omega - g_1^{T-1}. \) The corresponding cases are (ii.a) and (ii.c). Agent 1’s deviation benefit in (ii.a) is:

\[
D_1 = \omega + (a-1)(g_1^{T-1} + \Delta) + ag_2^{T-1} + \omega + a(g_2^T - g_1^T + \Delta) \\
= [\omega + (a-1)g_1^{T-1} + ag_2^{T-1} + \omega + (a-1)g_1^T + ag_2^T] \\
= (2a-1)(\Delta - g_1^T) \\
= (2a-1)[\omega - (g_1^{T-1} + g_1^T)] > 0. \tag{B.1}
\]

\(^{13}\)Take (ii.a) for an example. \( \hat{g}^r \) ensures contribution equality since:

If: \( \sum_{r=1}^{k-1} g_1^r + g_1^k + \sum_{r=k+1}^{T} g_1^r = \sum_{t=r}^{k-1} g_2^r + g_2^k + \sum_{r=k+1}^{T} g_2^r, \)
then: \( \sum_{r=1}^{k-1} g_1^r + g_1^k + \Delta + (T-k-1)\omega = \sum_{t=r}^{k-1} g_2^r + g_2^k + (T-k-1)\omega + \sum_{r=k+1}^{T} g_2^r - \sum_{r=k+1}^{T} g_1^r + \Delta. \)

In contrast, (ii.c) and (ii.d) are two corner situations where contribution inequity may not be eliminated in an equilibrium following a deviation.
Income received in earlier rounds cancels out and does not appear in the calculation. If the applicable case is instead (ii.c), agent 1 has:

\[
D_1 = \omega + (a - 1)(g_1^{T-1} + \Delta) + ag_2^{T-1} + \omega + a\omega - \alpha \left[ (\sum_{t=1}^{T-1} g_1^t + \Delta) - (\sum_{t=1}^{T-1} g_2^t + \omega) \right] 
\]

\[
- \left[ \omega + (a - 1)g_1^{T-1} + ag_2^{T-1} + \omega + (a - 1)g_1^T + ag_2^T \right] 
\]

\[
= (a - 1)[\Delta - g_1^T] + a(\omega - g_2^T) - \alpha \left[ \sum_{t=1}^{T-1} g_1^t - \sum_{t=1}^{T-1} g_2^t + \Delta - \omega \right] \quad (B.2)
\]

Observe that in (B.2), \(D_1\) decreases in \(\Delta\). The maximum deviation benefit is obtained instead when \(\Delta\) approaches its lower bound in its range, i.e., \(\omega + g_1^T - g_2^T\). So \(D_1\) approaches the following from above:

\[
(2a - 1)(\omega - g_2^T) - \alpha \left[ \sum_{t=1}^{T} g_1^t - \sum_{t=1}^{T} g_2^t \right] = (2a - 1)(\omega - g_2^T) > 0. \quad (B.3)
\]

Therefore \(D_1\) is also positive in the case of (ii.c). These positive deviation benefits form the contradiction needed to support the Lemma when \(k = T - 1\).

Now suppose that all rounds after \(k\) satisfy the minimum aggregate contribution condition. We show it is then also the case for round \(k\). Again use contradiction and suppose \(g_i^t\) is an equilibrium with:

\[
\sum_{r=t}^{T} g_i^r \geq (T - t)\omega, \forall t \in [k + 1, T], i = 1, 2,
\]

\[
\sum_{r=k}^{T} g_i^r < (T - k)\omega \quad (B.4)
\]

Agent 1 can deviate with the maximum \(\Delta > 0\) in round \(k\) (i.e., \(\Delta = \omega - g_1^k\)) to
receive according to (ii.a):

\[
D_1 = \omega + (a - 1)(g_1^k + \Delta) + ag_2^k + (T - k)\omega +
+ (a - 1)(T - k - 1)\omega + a[(T - k - 1)\omega + G^{k+1}]
- [\omega + (a - 1)g_1^k + ag_2^k + (T - k)\omega + (a - 1)\sum_{r=k+1}^T g_1^r + a\sum_{r=k+1}^T g_2^r]
= (2a - 1)[\Delta + (T - k - 1)\omega - \sum_{r=k+1}^T g_1^r]
= (2a - 1)[(T - k)\omega - \sum_{r=k}^T g_1^r] > 0, \quad (B.5)
\]

or according to (ii.c) to receive a strictly larger value than:

\[
(2a - 1)[(T - k)\omega - \sum_{r=k}^T g_1^r] > 0 \quad (B.6)
\]

through a similar procedure of (B.2). Consequently, agent 1 will be tempted to deviate in round \( k \), which contradicts \( g_1^t \) being an equilibrium.

We hence have proved the Lemma.

\[\square\]

**Proposition 2.** For the dynamic public goods game played between two inequity-averse agents, given \( a + \beta - 1 > 0 \), a pure strategy subgame perfect equilibrium, \( (g_1^t, g_2^t) \), is fully characterized by:

(a) \( \sum_{t=1}^T g_1^t = \sum_{t=1}^T g_2^t \),

(b) \( \sum_{r=t}^T g_i^r \geq (T - t)\omega, \forall t \in [1, T], i = 1, 2. \)

*Proof.* As Lemma 1 and 2 have shown that there is no equilibrium that does not comply with (a) and (b), we only need to prove that an equilibrium described above is immune to deviations. We still refer to part (ii) of Lemma 2 for the off-equilibrium strategies. (ii.a), (ii.b) and (ii.d) are the relevant scenarios in the current equilibrium.\(^{14}\) Lemma 2 has shown that (ii.a) is immune to

\(^{14}\)(ii.c) is irrelevant because it requires at least \( \sum_{r=k+1}^T g_2^r - \sum_{r=k+1}^T g_1^r + \omega - g_1^k > \omega \), or equivalently \( \sum_{r=k+1}^T g_2^r > \sum_{r=k}^T g_1^r \). The proposed equilibrium however has \( \sum_{r=k+1}^T g_2^r \in [(T - k - 1)\omega, (T - k)\omega] \) and \( \sum_{r=k}^T g_1^r \in [(T - k)\omega, (T - k + 1)\omega] \). Hence (ii.c) is not relevant.
deviations as long as $\sum_{t=t}^{T} g_t^r \geq (T-t)\omega$. Through a similar procedure as in (B.5), the case of (ii.b) can be shown as free of deviation, too, for deviation benefits can be obtained as negative as:

$$D_1 = (2a - 1)[(T - k - 1)\omega - \sum_{r=k+1}^{T} g_r^2] \leq 0$$

The case of (ii.d) is actually similar to that of (ii.c), except that now agent 1 may end up with an advantageous inequity (despite her willingness to catch up in contributions) such that $\beta$ applies to the psychological cost. Accordingly, the largest possible $\Delta$ in this scenario will bring the highest deviation return, which approaches from below:

$$(2a - 1)[(T - k - 1)\omega - \sum_{r=k+1}^{T} g_r^2] \leq 0$$

The equilibrium is therefore always immune to deviations. At the same time, an equilibrium defined by the proposition is also subgame perfect, for in each scenario discussed in above, part (a) is conformed with to the highest possible level (contribution equity in (ii.a) and (ii.b), and minimal inequity in (ii.d)), meanwhile part (b) is also respected.

We hence complete the proof for the Proposition.

The Limited Punishment for Deviations (LPD) principle says:

Whenever possible, $P_k^1(\Delta) \in [0, |\Delta|]$, with $|\Delta|$ for $\Delta$’s absolute value.

where $P_k^1(\Delta)$ is the punishment for agent 1’s deviation on round $k$ from the equilibrium. Notice that in Proposition 2, $P_k^1$ can be as high as $\sum_{r=k+1}^{T} g_r^1 - (T-k-1)\omega$ in (ii.a) and $\sum_{r=k+1}^{T} g_r^2 + \Delta - (T-k-1)\omega$ in (ii.b), both are larger than $|\Delta|$. With LPD restricting the harshness of punishment, Proposition 3

\footnote{To see this for (ii.a), use the largest possible $\Delta = \omega - g_k^1$. $P_k^1 - \Delta = \sum_{r=k+1}^{T} g_r^1 - (T-k-1)\omega - \omega + g_k^1 = \sum_{r=k}^{T} g_r^1 - (T-k)\omega$, which is positive in an equilibrium. The case for (ii.b) is more obvious.}
reduces the equilibrium set in Proposition 2.

**Proposition 3.** Under LPD, a pure strategy equilibrium of the dynamic game played between two inequity-averse agents is featured with the following contributions:

(a) \( g_1^t = g_2^t, \forall t; \)

(b) \( g_1^t = g_2^t = \omega, \forall t \) with at most one exceptional \( k \in [1, T], \) when \( g_1^k = g_2^k \in [0, \omega]. \)

**Proof.** Start with an equilibrium that conforms to Proposition 2, which has more than one round in which not both players contribute \( \omega \), i.e., (b) of Proposition 3 is not satisfied. Without loss of generality, let \( g_1^k < \omega \) and \( \sum_{r \neq k} g_2^r < (T - 1)\omega \). We look at an equilibrium strategy that follows (ii) in Lemma 2 but employs punishment according to LPD. That is:

* \( \sum_{t=1}^{T} g_1^t = \sum_{t=1}^{T} g_2^t. \)

* If \( \hat{g}_i = (g_i^k + \Delta, g_i^k), \forall \Delta \in [-g_1^k, \omega - g_1^k], \) then \( \hat{g}_i = (\hat{g}_i^r, \hat{g}_i^s), \) with \( \sum_{t=r}^{T} \hat{g}_i^t \geq (T - t)\omega, i = 1, 2, \forall t \in [k + 1, T], \) and:

(3.a) \( \sum_{r=k+1}^{T} \hat{g}_1^r = \sum_{r=k+1}^{T} g_1^r - P_1^{k+1}(\Delta) \) and \( \sum_{r=k+1}^{T} \hat{g}_2^r = \sum_{r=k+1}^{T} g_2^r + \Delta - P_1^{k+1}(\Delta), \) where \( P_1^{k+1}(\Delta) \in [0, |\Delta|], \) if \( \sum_{r=k+1}^{T} g_2^r + \Delta \in [(T - k - 1)\omega, (T - k)\omega]; \)

(3.b) \( \sum_{r=k+1}^{T} \hat{g}_1^r = \sum_{r=k+1}^{T} g_1^r - H_2^{k+1}(\Delta) - P_1^{k+1}(\Delta) \) and \( \sum_{r=k+1}^{T} \hat{g}_2^r = \sum_{r=k+1}^{T} g_2^r + \Delta - H_2^{k+1}(\Delta) - P_1^{k+1}(\Delta), \) where \( P_1^{k+1}(\Delta) \in [0, \sum_{r=k+1}^{T} g_1^r - H_2^{k+1}(\Delta) - (T - k - 1)\omega], \) if \( \sum_{r=k+1}^{T} g_2^r + \Delta > (T - k)\omega \) and \( \sum_{r=k+1}^{T} g_1^r - H_2^{k+1}(\Delta) - (T - k - 1)\omega \geq 0; \)

(3.c) \( \sum_{r=t}^{T} \hat{g}_1^r = (T - t)\omega \) and \( \sum_{r=t}^{T} \hat{g}_2^r = (T - t + 1)\omega, \forall t \in [k + 1, T], \) if \( \sum_{r=k+1}^{T} g_2^r + \Delta > (T - k)\omega \) and \( \sum_{r=k+1}^{T} g_1^r - H_2^{k+1}(\Delta) - (T - k - 1)\omega < 0; \)

(3.d) \( \sum_{r=t}^{T} \hat{g}_1^r = (T - t + 1)\omega, \sum_{r=t}^{T} \hat{g}_2^r = (T - t)\omega, \forall t \in [k + 1, T], \) if \( \sum_{r=k+1}^{T} g_2^r + \Delta < (T - k - 1)\omega. \)
(3.a) and (3.b) cover situations corresponding to scenarios (ii.a) and (ii.b) in Lemma 2, with (3.b) for a special case of (ii.a). (3.c) and (3.d) correspond to (ii.c) and (ii.d), respectively. Although (3.c) is still irrelevant as in Proposition 2 and (3.d) keeps making any deviation unprofitable, agent 1 can choose $\Delta > 0$ so that according to (3.a), she gains a deviation benefit of:

$$D_1 = (2a - 1)(\Delta - P_1^k(\Delta)) > 0$$

This profit disappears only when $\sum_{r=k+1}^{T} g_2^r = (T - k)\omega$, so that no positive $\Delta$ triggers scenario (3.a). It is then easy to show that, in that case, $D_1$ is non positive under (3.b) as well.

This indicates that if $g_1^k < \omega$, then $g_2^T = \omega, \forall t \in [k + 1, T]$. Combining it with the equal aggregate contribution condition, we reach both (a) and (b) in the Proposition.

B.2 Prove Proposition 4

Proposition 4 depicts the cooperation between an inequity-averse agent and a selfish agent, when type information is known in the dynamic game. Use $IA$ to represent an inequity-averse agent and $S$ for a selfish one, we prove the Proposition in below.

Proof. First, $\sum_{t=1}^{T} g_{IA}^t = \sum_{t=1}^{T} g_S^t$ can be shown in a similar way as in Lemma 1. Under assumption 1, $IA$ always matches the other’s action.

Second, in any equilibrium, $g_S^T = 0$, and:

- either $g_{IA}^T = \omega$,
- or $g_{S}^{T-1} = \omega, g_{IA}^T \in [\frac{1-a}{a} \omega, \omega]$.

Suppose not. If $\max\{g_{IA}^T, g_{S}^{T-1}\} < \omega$, let $S$ deviate by contributing $g_{S}^{T-1} + \delta$ in round $T - 1$, with $\delta > 0$. $IA$ will reduce the inequity by choosing a contribution level close enough to $g_{IA}^T + \delta$ in round-$T$. This deviation
brings $S$ a net gain of:

$$D_S = \left[ \omega + (a - 1)(g_s^{T-1} + \delta) + a g_{IA}^{T-1} + \omega + a(g_{IA}^T + \delta) \right] - \left[ \omega - (1 - a)g_s^{T-1} + a g_{IA}^{T-1} + \omega + a g_{IA}^T \right] = (2a - 1)\delta.$$ 

In the other case, if $g_s^{T-1} = \omega$ but $g_{IA}^T < \frac{1-a}{a} \omega$, let $S$ deviate to 0 in round $T - 1$. $IA$ will then match with $\hat{g}_{IA}^T = 0$. The deviation is then profitable for $S$, as for $g_{IA}^T < 1 - a a \omega$:

$$D_S = \left[ \omega + ag_{IA}^{T-1} + \omega \right] - \left[ \omega - (1 - a)\omega + ag_{IA}^{T-1} + \omega + ag_{IA}^T \right] = (1 - a)\omega - ag_{IA}^T > 0.$$ 

**Third,** without imposing LPD, $\sum_{t=1}^{T-1} g_i^t \geq (T - 1 - t)\omega, i = IA, S$, and

- either $g_{IA}^T = \omega$,
- or $g_s^{T-1} = \omega, g_{IA}^T \in [\frac{1-a}{a} \omega, \omega].$

The proof is basically the same as for proposition 2. With a high enough final-round equilibrium contribution by $IA$, $S$ is behaving as if also being averse to inequity in all early rounds.

**Finally,** with LPD, $S$ contributes all the time $\omega$ till round $T - 1$, and $IA$ fulfills $g_{IA}^T \geq \frac{1-a}{a} \omega$ in addition to equal aggregate contributions. To show this, use contradictions and suppose an equilibrium profile $(g_{IA}^t, g_s^t)$ with at least one round $k < T$ in which $g_s^k < \omega$.

- if $g_{IA}^r = \omega, \forall r > k$, there must be at least one round $k' < k$ with $g_{IA}^{k'} < \omega$ to ensure equal aggregate contributions. Then $IA$ can deviate in round $k'$ to $\hat{g}_{IA}^{k'} = g_{IA}^{k'} + \delta$, with $\delta > 0$. One equilibrium following this can be $\hat{g}_s^t = g_s^t, \forall t \neq k'$, and $\hat{g}_s^k = g_s^k + \delta - P_{IA}(\delta)$; meanwhile $\hat{g}_{IA}^t = g_{IA}^t, \forall t \neq k''$, and $\hat{g}_{IA}^{k''} = g_{IA}^{k''} - P_{IA}(\delta)$. $k''$ is a random round after $k'$. With $P \leq \delta$, that makes a non-negative deviation benefit for $IA$.

---

16Technically, it is necessary to notice that, in any subgame perfect strategy, $IA$’s round-$T$ contribution after a deviation is $\hat{g}_{IA}^T = \omega$. 

37
- if instead $\exists k'' > k$ where $g_{IA}^{k''} < \omega$, $S$ can again make the $\delta$-deviation and benefit from $IA$’s contribution match in this round $k''$.

\[ \square \]

### B.3 Prove Proposition 5

**Proof.** To start, we assume that if $IA$ needs to respond to the other’s action in certain round $k$, she chooses the nearest future round $t > k$ that can accommodate this action.

(a) First we prove that when $1 - p < \min\left[\frac{1-a}{a}, \frac{1-a+a+\alpha}{\alpha+\beta}\right]$, there is a unique equilibrium with agents pooling on $g^t = (0, 0), \forall t$:

- $g^t = (0, 0)$ is the only pooling equilibrium, because:

  There exists no pooling equilibrium with positive contributions:

  Suppose there is one. Since in a pooling equilibrium round $T$ is like a static game in terms of belief, (b) of Proposition 1 requires $g^T = (0, 0)$. Let round-$k$ be the last round with positive contributions. A selfish agent will always choose 0 in this round as there is nothing to gain in the future, which contradicts the equilibrium being with positive contributions.

  $g^t = (0, 0), \forall t < T$ is not subject to deviations, for deviations bring negative profits. Take a selfish agent as an example, even though with probability $1 - p$ there is an $IA$ who matches his deviation, the gain is:

  \[
  D_s = T \cdot \omega + (a - 1)\omega + a(1 - p)\omega - T \cdot \omega
  \]

  - There exists no separating equilibrium. For if there was one, there must be $g_{IA}^T \leq \frac{1-a}{a} \frac{1}{1-p} g_{IA}^{T-1}$ for the revelation to happen in round $T - 1$, which will be shown a bit later in our proof. An $IA$ could then deviate in $T - 1$
by mimicking an $S$ to have:

$$D = [\omega + a(1-p)g_{IA}^{T-1} + \omega + (1-p)(a-1)g_{IA}^{T-1}] -$$

$$- [\omega + (a-1)g_{IA}^{T-1} + a(1-p)g_{IA}^{T-1} + \omega + (2a-1)(1-p)g_{IA}^{T} - \alpha pg_{IA}^{T-1}]$$

$$= (1-a+\alpha)pg_{IA}^{T-1} - (2a-1)(1-p)g_{IA}^{T}$$

$$\geq (1-a+\alpha)pg_{IA}^{T-1} - (2a-1)\frac{1-a}{a}g_{IA}^{T-1} > 0, \text{ for } 1-p < \frac{1-a}{a}$$

Where the last line uses $g_{IA}^{T} \leq \frac{1-a}{a} \frac{1}{1-p}g_{IA}^{T-1}$.

(b) Now we prove that when $1-p \geq \max \left[ \frac{1-a+\alpha}{a+\beta}, \frac{1-a+\alpha}{a+\alpha} \right]$, there are only two types of pooling equilibrium. The one with $(\omega, \omega)$ till the penultimate round and then $g_{IA}^{T} = x \geq \frac{1-a}{a} \frac{\omega}{1-p}$ is proved in below. The other that has at most one round when $g_{IA}^{k} < \omega$ is actually shown during the proof.

- This is an equilibrium because:

  - An IA does not deviate in round $T$. Any $\hat{g}_{IA}^{T} = x - \delta, \forall \delta > 0$ makes no profit:

    $$D_{IA} = [\omega + (a-1)\hat{g}_{IA}^{T} + a(1-p)x - \beta(1-p)(x - \hat{g}_{IA}^{T}) - \alpha p\hat{g}_{IA}^{T}] -$$

    $$- [\omega + (a-1)x + a(1-p)x - \alpha px]$$

    $$= (1-a+\alpha)\delta - (\alpha + \beta)(1-p)\delta$$

    $$\leq 0, \text{ for } 1-p \geq \frac{1-a+\alpha}{\alpha + \beta}$$

  - Suppose a deviation of $g_{IA}^{k} = \omega - \delta$, with $\delta \in [0, \omega], \forall k < T$ leads to no belief changes concerning deviator’s type but a punishment $P \in [0, \delta]$ in round $k+1$. It is easy to show that the deviation profit is negative for both types if $k+1 < T$. For $k+1 = T$, the deviation
profits are:

\[ D_{IA} = \left[ \omega + (a - 1)(\omega - \delta) + a\omega + \omega + (a - 1)(x - P) + a(1 - p)(x - P - \delta) - \alpha p(x - P) \right] - \]

\[ - \left[ \omega + (a - 1)\omega + a\omega + \omega + (a - 1)x + a(1 - p)x - \alpha px \right] \]

\[ = \left[ (1 - a) - a(1 - p) \right] \delta + \left[ (1 - a + \alpha) - (a + \alpha)(1 - p) \right] P \]

\[ \leq 0, \text{ for } 1 - p \geq \frac{1 - a + \alpha}{a + \alpha} > \frac{1 - a}{a}, \text{ and} \]

\[ D_S = \left[ \omega + (a - 1)(\omega - \delta) + a\omega + \omega + a(1 - p)(x - P - \delta) \right] - \]

\[ - \left[ \omega + (a - 1)\omega + a\omega + \omega + a(1 - p)x \right] \]

\[ = \left[ (1 - a) - a(1 - p) \right] \delta - a(1 - p)P \]

\[ \leq 0. \]

We skip to expand the corner situation where a player drops contribution to 0 in round \( T - 1 \). That also gives a negative benefit since \( \hat{g}_T^{IA} = 0 \) follows in the next round.

- Other pooling equilibria with more than one round (including round \( T \)) where \( g^t = (g^t_{IA}, g^t_S) < \omega \) are prone to deviations. To show this, take an example equilibrium with \( g^t_{IA} = g^t_S = \omega, \forall t < T \) and \( t \neq k; \)

\( g^k_{IA} = g^k_S = z < \omega; g^T_{IA} = x, g^T_S = 0 \). After having a deviation to \( z + \delta \) in round-\( k \), since rounds between \( k + 1 \) and \( T \) will not accommodate \( g^t_i = \omega - P + \delta \) with \( P < \delta \), we rely on round \( T \) to realize the deviation matching and punishment. We also require that following the deviation:

- the belief stays unchanged and players adopt \( P \in \left[ 0, \frac{a(1 - p) - (1 - a)\delta}{a(1 - p)} \right] \); or

- the deviator is believed to be inequity-averse and players adopt \( P \in \left( \frac{a(1 - p) - (1 - a)\delta}{a(1 - p)} \delta, \delta \right) \).

The game played after the deviation is then nearly unchanged to the
equilibrium path. The deviation benefits are:

\[ D_{IA} = [\omega + (a - 1)(z + \delta) + az + (T - k - 1)2\omega + \omega + (a - 1)(x - P) + a(1 - p)(x - P + \delta) - \alpha p(x - P + \delta)] - \]
\[ - [\omega + (2a - 1)z + (T - k - 1)2\omega + \omega + (a - 1)x + a(1 - p)x - \alpha px] \]
\[ = [(a - \alpha - 1) + (a + \alpha)(1 - p)](\delta - P) \geq 0, \text{ for } 1 - p \geq \frac{1 - a + \alpha}{a + \alpha}, P < \delta; \]

\[ D_s = [\omega + (a - 1)(z + \delta) + az + (T - k - 1)2\omega + \omega + a(1 - p)(x - P + \delta)] - \]
\[ - [\omega + (2a - 1)z + (T - k - 1)2\omega + \omega + a(1 - p)x] \]
\[ = (a - 1)\delta + a(1 - p)(\delta - P) > 0 \text{ for } 1 - p > \frac{1 - a}{a}, P \leq \frac{a(1 - p) - (1 - a)}{a(1 - p)} \delta. \]
\[ \leq 0 \text{ for } P \geq \frac{a(1 - p) - (1 - a)}{a(1 - p)} \delta. \]

So if \( z \) and \( x \) are both smaller than \( \omega \), agents do deviate and beliefs are updated correctly. This forms a contradiction to the equilibrium. Besides, in the case that \( x = \omega \), an IA becomes indifferent between deviating and not. That is then exactly the second kind of pooling equilibrium that we have proposed.

Note that the belief may also be updated in a different way, for example a deviator is always considered selfish. The game after the deviation will of course be played differently. We do find this belief to be consistent in the way that either only the selfish agent deviates, or neither agent deviates at all. However, treating every deviator as selfish is comparable to imposing the harshest punishment for deviations. Since we prefer LPD, we shall be equally careful to adopt a belief system like this.

(c) Now prove that when \( 1 - p \geq \frac{(1 - a)^2 + \alpha a}{a(1 - a) + \alpha a} \), there exists one type of separating equilibrium with both types contributing fully until \( T - 2 \) round.
• The proposed contributions can be realized in an equilibrium:

  - Easy to show that the final round contribution \( y \) is small enough to prevent a selfish agent from mimicking an inequity-averse agent, meanwhile is large enough to prevent \( IA \) from mimicking \( S \) in round \( T - 1 \).

  - To show that there is no \( \delta \)-deviations in any round \( t \), suppose beliefs are not updated after \( \tilde{g}_t^I = \omega - \Delta \). Also suppose the strategy after deviation involves matching and punishment in round \( t + 1 \), after which contributions return to the original equilibrium path. I show below that a deviation at \( t = T - 1 \) is not profitable. The cases with \( t < T - 1 \) are more obvious.

For \( 1 - p \geq \frac{1-a+\alpha}{\alpha+\beta} \), deviation benefits are:

\[
D_{IA} = \left\{ \omega + (a - 1)(\omega - \Delta) + a(1 - p)\omega + \omega + (1 - p)\right\} \\
\left\{[(a - 1)(y - P) + a(y - P - \Delta)] - \alpha p(\omega - \Delta)\right\} - \\
\left[\omega + (a - 1)\omega + a(1 - p)\omega + \omega + (1 - p)(2a - 1)y - \alpha p\omega\right]
\]

\[
= \left[(1 - a + \alpha) - (a + \alpha)(1 - p)\right] \Delta + \left[(1 - 2a)(1 - p)\right] P
\]

\[
\leq 0
\]

\[
D_S = \left[\omega + (a - 1)(\omega - \Delta) + a(1 - p)\omega + \omega + (1 - p)a(y - P - \Delta)\right] - \\
\left[\omega + a(1 - p)\omega + \omega\right]
\]

\[
= (a - 1)(\omega - \Delta) + a(1 - p)(y - \Delta) - a(1 - p)P
\]

\[
< 0 \text{ for } y \leq \frac{1-a}{a} \frac{\omega}{1-p}
\]

For \( 1 - p < \frac{1-a+\alpha}{\alpha+\beta} \), these negative deviation profits will not change.

• The proposed separating equilibrium is not unique. For example, suppose in an equilibrium \( g_{IA}^t = g_s^t = \omega, \forall t < T-1, t \neq k; g_{IA}^k = g_s^k = z < \omega; g_{IA}^{T-1} = g_s^T = 0, g_{IA}^{T-1} = \omega, g_{IA}^T = y \) if \( g^{T-1} = (\omega, \omega) \) and 0 otherwise. Following (b), let \( \Delta \)-deviations be matched and punished only in round \( T \), and the belief does not update, so separation is still in round \( T - 1 \). For
an IA agent, the profit of deviating in round \( k \) is:

\[
D_{IA} = \omega + (2a - 1)z + (a - 1)\Delta + (T - k - 2)2a\omega + \omega + (a - 1)\omega + a(1 - p)\omega + \omega + (1 - p)\left[(a - 1)(y - P) + a(y - P + \Delta)\right] - \alpha p(\omega + \Delta) - \left[\omega + (2a - 1)z + (T - k - 2)2a\omega + \omega + (a - 1)\omega + a(1 - p)\omega + \omega + (1 - p)(2a - 1)y - \alpha p\omega\right]
\]

\[
= \left[-(1 - a + \alpha) + (a + \alpha)(1 - p)\right]\Delta - (2a - 1)(1 - p)P
\]

This is positive only if \( P < \frac{(a + \alpha)(1 - p) - (1 - a + \alpha)(1 - p)}{(2a - 1)(1 - p)}\). So under LPD, this equilibrium may just survive deviations. This result is also robust to \( k = T - 1 \). However, in any equilibrium the less-than-maximum-contribution round \( k < T \) must be unique, otherwise profitable deviation opportunities arise.

- Finally, we show that separation can only happen at round \( T - 1 \). Suppose in an equilibrium types are separated in round \( k < T - 2 \) with \( g_{IA}^k = z \leq \omega \) and \( g_s^k = 0 \). A selfish agent would be able to mimic the inequity-averse agent in this round for a deviation benefit of:

\[
D_s = \left\{\omega + (a - 1)z + a(1 - p)z + (1 - p)\left[(T - k - 1)2a\omega + \omega + ax\right] + p\left[\omega + az + (T - k - 2)2a\omega + \omega + 2(1 - a)\omega\right]\right\} - \left\{\omega + a(1 - p)z + (1 - p)\left[\omega + (a - 1)z + (T - k - 2)2a\omega + \omega\right] + p\left[(T - k)\omega\right]\right\}
\]

\[
= (2a - 1)[pz + p(T - k - 2)\omega + (1 - p)\omega] + 2p(1 - a)\omega + a(1 - p)x
\]

\[
> 0
\]

What we have is, after the revelation, the game is played under complete information. Specifically, if it is revealed that one agent is selfish and the other is inequity-averse, the selfish one can catch up by playing \( z \) in the next round, in the meantime the inequity-averse agent contributes 0. Then the two play a complete information game of \( T - k - 1 \) rounds following Proposition 4. In this case, a selfish agent by mimicking an inequity-averse agent can trick the other agent to cooperate after round \( k \). He is able to earn up to \( 2(1 - a)\omega \) more than a real inequity-averse
agent in material payoffs if the other agent is revealed to be selfish, because he will not contribute in round $T - 1$ and round $T$, and the highest contributions of an inequity-averse agent in these two rounds are both $\omega$. This explains the payoffs in (B.7).

The conclusion does not change for $t = T - 2$. □

### B.4 Elaborating the Reputation Equilibrium

In this section we provide details for the reputation equilibrium of Proposition 6. Although it is not a formal proof, the elaboration already shows the validity of this equilibrium. We will start backwardly from the final round.

Denote $\{\epsilon^t\}$ as a series of small positive numbers with values uniquely determined according to rules that we will shortly define. Denote an inequity-averse agent by $IA$, the selfish who contributes $\omega$ in round $k$ by $S_R$, and a selfish agent who contributes 0 in round $k$ by $S_{UR}$ (R for Revealed and UR for UnRevealed). Further denote the probability that $S_R$ contributes $\omega$ in round $t$ by $z^t$, with $t > k$. Denote the probability that an unknown-type agent contributes $\omega$ in round-$t$ by $q^t$, the belief about an unknown-type agent being $IA$ by $1 - p^t$, and the probability that $S_{UR}$ contributes $\omega$ in round-$t$ by $r^t$.

Suppose that if 0 is contributed by both agents in round $k$, or by either agent in a round following $k$, $(0,0)$ will be the contributions till the end of game. For any round $t$ mentioned below, we restrict attention to the situation with a history of $g^k = (\omega,0)$, and $g^\tau = (\omega,\omega), \forall k < \tau < t$. That is, after round $k$, the group is known to consist of one $S_R$, and the other either $IA$ or $S_{UR}$, and all the randomizations happen to have realized contributions of $\omega$. We argue that:

- $IA$ is willing to contribute $\omega$ in each round $t > k$, provided $S_R$ contributes $\omega$ in the previous round. In this way she keeps equal contribution with the other meanwhile attains a highest overall payoff in the equilibrium.

- Since $S_R$ contributes 0 in round $T$, $S_{UR}$ will also contribute 0 in round $T - 1$.  

44
In round $T - 1$, $S_R$ contributes $\omega$ for sure if:

$$\omega + (a - 1)\omega + aq^{T-1}\omega + \omega + aq^{T-1}\omega$$

$$> \omega + aq^{T-1}\omega + \omega$$

$$\Rightarrow q^{T-1} > \frac{1 - a}{a}.$$

Let $q^{T-1} = \frac{1-a}{a} + \epsilon^{T-1}$.

- In round $T - 2$, $S_{UR}$ randomizes between contributing $\omega$ and 0 if:

$$\omega + (a - 1)\omega + az^{T-2}\omega + 2\omega + az^{T-2}\omega$$

$$= \omega + az^{T-2}\omega + 2\omega$$

$$\Rightarrow z^{T-2} = \frac{1 - a}{a}$$

This means $S_R$ is randomizing between contributing $\omega$ and 0 in this round, which requires:

$$\omega + (a - 1)\omega + aq^{T-2}\omega + q^{T-2}V_R^{T-1} + (1 - q^{T-2})2\omega$$

$$= \omega + aq^{T-2}\omega + 2\omega$$

$$\Rightarrow q^{T-2}q^{T-1} = \frac{1 - a + q^{T-2}}{2},$$

where $V_R^{T-1}$ represents $S_R$’s continuity utility in round $T - 1$ after $g^{T-2} = (\omega, \omega)$, which is just the left hand side of (B.8).

- In round $T - 3$, $S_{UR}$ randomizes if:

$$\omega + (a - 1)\omega + az^{T-3}\omega + z^{T-3}V_{UR}^{T-2} + (1 - z^{T-3})3\omega$$

$$= \omega + az^{T-3}\omega + 3\omega$$

$$\Rightarrow z^{T-3}z^{T-2} = \frac{1 - a}{a}.$$

Similarly, $V_{UR}^{T-2}$ represents $S_{UR}$’s continuity utility in round $T - 2$ following $g^{T-3} = (\omega, \omega)$. Since $z^{T-2} = \frac{1-a}{a}$, there must be $z^{T-3} = 1$, which corresponds to $S_R$ contributing $\omega$ for sure in this round. This requires

$$\omega + (a - 1)\omega + aq^{T-3}\omega + q^{T-3}V_R^{T-2} + (1 - q^{T-3})3\omega$$

$$> \omega + aq^{T-3}\omega + 3\omega$$

$$\Rightarrow q^{T-3}q^{T-2} > \frac{1 - a}{a}.$$
Let $q^{T-3}q^{T-2} = \frac{1-a}{a} + \epsilon^{T-3}$.

- In round $T - 4$, $S_{UR}$ randomizes if:

\[
\omega + (a - 1)\omega + az^{T-4}\omega + z^{T-4}V_{UR}^{T-3} + (1 - z^{T-4})4\omega \\
= \omega + az^{T-4}\omega + 4\omega \\
\Rightarrow \quad z^{T-4} = \frac{1-a}{a},
\]

where $z^{T-3} = 1$ is applied in the last line. It corresponds to $S_R$ randomizing in this round:

\[
\omega + (a - 1)\omega + aq^{T-4}\omega + q^{T-4}V_{R}^{T-2} + (1 - q^{T-4})4\omega \\
= \omega + aq^{T-4}\omega + 4\omega \\
\Rightarrow \quad q^{T-4}q^{T-3} = \frac{1-a}{a} \frac{1+q^{T-2}}{1+q^{T-2}}.
\]

- Rolling this analysis backward, we have the following systems, in which $n \in \mathbb{N}$:

\[
\{z^t\} \quad z^{T-(2n-1)} = 1 \quad \{q^t\} \quad q^T = 1 \\
\quad z^{T-2n} = \frac{1-a}{a} \quad \quad \quad q^{T-(2n-1)}q^{T-(2n-2)} = \frac{1-a}{a} + \epsilon^{T-(2n-1)} \\
\quad \quad \quad q^{T-2n}q^{T-(2n-1)} = \frac{1-a}{a} \frac{1+q^{T-2n}}{1+q^{T-(2n-2)}}
\]

Here every $z^t$ is uniquely defined, so is every $q^t$ given the set $\{\epsilon^t\}$.

- Now suppose $k = T - (2n - 1)$. In round $k$, an $S$ randomizes between contributing $\omega$ to become $S_{R}$, and contributing 0 to become $S_{UR}$. Denote the probability that an agent contributes $\omega$ in this round by $\delta^k$. The selfish agent’s problem is:

\[
\omega + (a - 1)\omega + a\delta^k\omega + \delta^k(2n - 1)\omega + (1 - \delta^k)\left[V_{R}^{k+1}\right] \\
= \omega + a\delta^k\omega + \delta^k\left[V_{UR}^{k+1}\right] + (1 - \delta^k)(2n - 1)\omega.
\]

(B.9)

Meanwhile an inequity-averse agent contributes 0 for certain:

\[
\omega + (a - 1)\omega + a\delta^k\omega + \delta^k(2n - 1)\omega + \\
+(1 - \delta^k)[\omega + aq^{k+1}\omega + (2n - 2)\omega - \alpha(1 - q^{k+1})\omega] \\
\leq \omega + a\delta^k\omega + \delta^k\left[V_{IA}^{k+1}\right] + (1 - \delta^k)(2n - 1)\omega.
\]

(B.10)
In the LHS of (B.10), we have assumed that, whenever the situation is for $S_R$ to randomize, it is not attractive for $IA$ to contribute, due to the latter’s fear for disadvantageous inequity that associates with making contributions. $V_{IA}^{k+1}$ is $IA$’s continuity utility at round $k + 1$, which corresponds to matching the other’s action in the previous round. To give an example:

$$
V_{IA}^{T-2} = \omega + (a - 1)\omega + az^{T-2}\omega + z^{T-2}[\omega + (a - 1)\omega + az^{T-1}\omega + \\
+ z^{T-1}(\omega + (a - 1)\omega) + (1 - z^{T-1})(\omega)] + (1 - z^{T-2})2\omega,
$$
in which we have known that $z^{T-2} = \frac{1-a}{a}$ and $z^{T-1} = 1$.

The two conditions (B.9) and (B.10) can be reduced to:

$$
\begin{cases}
\delta^k & \leq \frac{\alpha(1-q^{k+1})}{\alpha(1-q^{k+1})+2(1-a)\left(\frac{1-a}{a}\right)(n-1)} \\
\delta^k & = \frac{aq^{k+1}(1-a)}{aq^{k+1}(1-a)}
\end{cases}
$$

(B.11)

where we have used $k = T - (2n - 1)$. A necessary condition for (B.11) is:

$$
q^{k+1} \leq \frac{\alpha + (1 - a)(\frac{1-a}{a})(n-1)}{\alpha + a(\frac{1-a}{a})(n-1)}
$$

One can show that this necessary condition can be satisfied by $\{q^t\}$ through carefully choosing $\{\epsilon^t\}$.

(B.11) then uniquely determines $\delta^k$ with $q^{k+1}$. The probability that a selfish agent contributes $\omega$ in this round is obtained as $r^k = \delta^k / p^k$. To have a feasible $r^k \in [0, 1]$, $p^k$ needs to be large, which is in line with our assumption that $1 - p$ is small in the current section.

- Throughout the game, the belief is updated according to the Bayes’ rule. That is:

$$
1 - p^{t+1} = \frac{1 - p^t}{q^t},
$$
The $T - 1$ round belief is then updated all the way from the prior belief as:

$$
1 - p^{T-1} = \frac{1 - p}{q^1q^2...q^{T-2}}
$$

(B.12)

We know $1 - p^{T-1} = q^{T-1} = \frac{1-a}{a} + \epsilon^{T-1}$. Also we know $q^t = 1, \forall t < k$. (B.12) will then uniquely define $n$ (equivalently, $k$) based on the prior belief $p$ and
the set \( \{ \epsilon^t \} \),

- Finally, with \( \{ q^t \} \) defined (based on \( \{ \epsilon^t \} \)) and the beliefs \( 1 - p^t \) obtained according to (B.12), \( S_{UR} \)'s strategy in terms of probabilities of contributing \( \omega \) in each round, \( r^t \), is obtained using the relationship: \( q^t = (1 - p^t) + p^t \cdot r^t \).

We have hence completely elaborated the proposed mixed strategy equilibrium and at the same time validated its existence.