Co-Financing Agreements and Reciprocity: When 'No Deal' is a Good Deal*

Dooseok Jang[†], Amrish Patel[‡] and Martin Dufwenberg[§]

November 9, 2016

Abstract

Institutions for co-financing agreements often exist to encourage public good investment. Can such frameworks deliver maximal investment when agents are motivated by reciprocity? We demonstrate that indeed they can, but not in the way one might expect. If maximal investment is impossible in the absence of the institution and public good returns are high, then an agreement signed by all parties cannot lead to full investment. However, if all parties reject the co-financing agreement, then an informal deal to invest can lead to full investment. Agreement institutions may thus do more than just facilitate the signing of formal agreements; they may play a critical role in igniting informal cooperation underpinned by reciprocity.

Keywords: Co-financing agreements, informal agreements, public goods, reciprocity.

JEL codes: C72, D03, F53, H41

*This paper extends and replaces Jang (2015). We thank Scott Barrett, Oana Borcan, Gautam Gowrisankaran, Bård Harstad, Doruk Iris, Kiryl Khalmetski, Georg Kirchsteiger, Ashley Langer, Derek Lemoine, David Reinstein, Stanley Reynolds, Alexander Sebald, Mark Stegeman and numerous seminar audiences for useful comments or discussions. Funding is gratefully acknowledged from the Jan Wallander and Tom Hedelius Foundation (P2012-0097-1).

[†]Korea Advanced Institute of Science and Technology; jangds@kaist.ac.kr

[‡]University of East Anglia; amrish.patel@uea.ac.uk

[§]University of Arizona, University of Gothenburg, CESifo; martind@eller.arizona.edu

1 Introduction

Institutions play an important role in creating the conditions for investment in public goods. Among other things, they facilitate the negotiation and enforcement of binding agreements. One common type of agreement is a co-financing, or cost-sharing, agreement; signatories make a binding commitment to co-finance each other's future public good investments. The agreement does not commit a signatory to invest in public goods per se. However, should any signatory initiate a public good investment, its co-signatories are committed to share the cost. Such agreements have been used to finance critical investment in public goods, ranging from disease eradication to climate change mitigation.[1]

Theoretically, co-financing agreements can increase public good investment (cf. Varian 1994).[2] This is because a signatory can be pivotal in inducing other signatories to invest in public goods, as only with its participation would the private cost of a public good be less than the private benefit. However, full investment remains impossible as if there are many signatories, an individual signatory is no longer pivotal, thus it deviates to not signing and not investing.

These insights rely on the assumption that agents care only about their material payoffs. Yet behaviour in public good contexts often exhibits conditional cooperation (e.g. Fischbacher et al. 2001), cooperating only if others do. Such behaviour can be rationalised using reciprocity theory (Rabin 1993, Dufwenberg and Kirchsteiger 2004 (D&K), Falk and Fischbacher 2006). It describes agents as having a desire to be kind to those who are kind to them, and unkind to those who are unkind to them. For example, if agent A invests, agent B may view A as kind and invest himself.

The implications of reciprocity for public good provision are both well established (e.g. Sugden 1984) and straightforward. If agents care enough

[1]In April 2016 The World Bank and The Asian Infrastructure Investment Bank signed a co-financing agreement focusing on water, transport and energy. Each party contributed $216 million to the first project, upgrading slums in Indonesia: www.brettonwoodsproject.org/2016/06/world-bank-and-aiib-signs-joint-co-financing-agreement. Such agreements are also signed by private companies. The Asian Development Bank, for instance, has an agreement with Chevron to invest in IT, construction and engineering education: www.adb.org/site/cofinancing/partners. One area where cost-sharing agreements are extensively used is in R&D investments (Katz 1986).

[2]Indeed higher investment is observed in related experimental games (Andreoni and Varian 1999, Falkinger et al. 2000 and Charness et al. 2007.)

about reciprocity, maximal investment is possible, otherwise it is not. By contrast, little is known about the implications of reciprocity for agreements over public goods.

An obvious question follows: How do co-financing agreements perform under reciprocity? More specifically, one may wonder: Can a co-financing mechanism deliver full investment? Does such investment follow if all players sign the agreement? Is it impossible if no-one signs the agreement? To answer these questions, we apply D&K's model of reciprocity to an agreements game where players choose whether or not to sign a cost-sharing agreement, then play a public good game. We find that if full investment was impossible with no mechanism and the public good return is high, then such investment remains impossible if all players sign. However, full investment *is* attainable if all players reject the agreement so that *no-one* signs.

For some intuition, consider the interaction of kindness and co-financing agreements. Roughly, D&K say that agent i is kind to agent j if i could have given j a much lower payoff by changing his behaviour. Agent i deviating from a situation where all players sign and invest does reduce j's payoff, but not by much, as the cost-sharing agreement still has many signatories thus provides large investment incentives. By contrast, if i deviates from a situation where no-one signs and all invest, j's payoff is reduced considerably as there is no such cost-sharing agreement. Kindness and hence reciprocity incentives to invest in public goods are thus larger when there are no signatories than when there are many.

Our results provide several important insights. First, the existence of an institution for making binding agreements is potentially critical for cooperation, even if an agreement is not actually signed. Second, while easy to overlook, agreements that are not signed may be the reason why we observe as many public goods as we do. Third, informal agreements, which in our model could be a legitimate interpretation of the "no-one signs but all invest anyway" outcome, may be able to achieve outcomes that formal agreements cannot. Fourth, and pointing to the more general feature underlying the previous insights, prior stages in games (here, an agreement stage) can increase a player's influence over others' payoffs, since others may condition their actions on his early choice. This increase in payoff influence, can "amplify" psychological payoffs (in our case kindness) and make otherwise impossible outcomes attainable.

We add to an active literature studying agreements (e.g. Barrett 2003, Harstad 2012, 2015, Battaglini and Harstad 2016, Martimort and Sand-

Zantman 2016) and an emerging literature on mechanism design where players have reciprocity preferences (Netzer and Volk 2014, Bierbrauer et al. 2015, Bierbrauer and Netzer 2016, Dufwenberg and Patel 2016).

Our particular mechanism, cost-sharing agreements, falls into a class of mechanisms where commitments on strategy-conditional side-payments are made before a game is played (Jackson and Wilkie 2005, Ellingsen and Paltseva 2016). Cost-sharing is an important case of models where agents make commitments to match others' public good investments (Guttman 1978, 1987, Boadway et al. 2007) or to compensate others for their investment (Varian 1994). Our game may also be relevant for agreements on R&D investment (Katz 1986) and International Environmental Agreements (IEAs) (Barrett 1994), if they involve binding co-financing.

Understanding the role of reciprocity in IEAs is important for environmental economists. Nyborg (2015) concurrently developed a model that extends D&K to cooperative games in order to apply it to Barrett's IEAs model. She finds that reciprocity can create weakly larger stable coalitions that exhibit higher abatement. Less closely related are Hadjiyiannis et al. (2012) and Kolstad (2014). The former studies the effect of a different notion of reciprocity on abatement in a two-player game with no possibility to sign an agreement. The latter examines the effect of equity- and efficiency-concerns (Charness and Rabin 2000) in an IEAs game.

We structure the paper as follows. Section 2 presents a set of preliminaries needed for our main result. Section 3 presents our main result on how full investment is impossible if all players sign, but is possible if no-one signs. Section 4 offers some further results before we conclude in Section 5.

2 Preliminaries

We introduce a public good game (2.1), an agreements game (2.2), and D&K's reciprocity model (2.3) which we apply to the public good game (2.4).

2.1 The public good game (Γ_P)

Let $N = \{1, \dots, n\}$ be the set of players where $n \geq 4$.[3] Each $i \in N$ simultaneously chooses $a_i \in \{0, 1\}$, where 1 corresponds to investing in a public

[3] We discuss $n < 4$ in Section 4.

good and 0 to not doing so. Let $a = (a_i)_{i \in N}$. Player i's material payoff is

$$\pi_i(a) = \beta \sum_{j \in N} a_j - \gamma a_i,$$

where β is the public good benefit and γ the cost. Assume that $n\beta > \gamma > \beta > 0$ so that the individual cost exceeds the benefit and all players investing maximises total payoffs. Γ_P has a unique NE where for all $i \in N$, $a_i = 0$.

2.2 The agreements game (Γ_A)

The agreements game appends a prior stage to the public good game where each player decides whether to sign a co-financing agreement: a binding commitment to share public good investment costs.

As before, $N = \{1, \ldots, n\}$, $n \geq 4$. In stage 1, the "sign-up stage", each $i \in N$ simultaneously chooses $a_i^1 \in \{0, 1\}$; 1 means signing the agreement, 0 not doing so. In stage 2, the "investment stage", each $i \in N$ simultaneously chooses $a_i^2 \in \{0, 1\}$; 1 means investing in the public good, 0 not doing so. Let $a^1 = (a_i^1)_{i \in N}$, $a^2 = (a_i^2)_{i \in N}$ and $a = (a^1, a^2)$.

Let $m(a^1) = \sum_{i \in N} a_i^1$ be the number of signatories, $x(a^2) = \sum_{i \in N} a_i^2$ the number of players who invest, $x^m(a) = \sum_{i \in N} a_i^1 a_i^2$ the number of signatories who invest. Player i's material payoff is

$$\pi_i(a) = \begin{cases} \beta x(a^2) - \gamma a_i^2 & \text{if } a_i^1 = 0 \\ \beta x(a^2) - \gamma \frac{x^m(a)}{m(a^1)} & \text{if } a_i^1 = 1 \end{cases},$$

where β is the public good benefit and γ the cost. The investment stage involves the same decision as in Γ_P, but with different payoff consequences for signatories. Signatories have made a binding commitment to share their investment costs. As before, $n\beta > \gamma > \beta > 0$ and to avoid knife-edge cases, assume $\frac{\gamma}{\beta}$ is non-integer and $n > \lceil \frac{\gamma}{\beta} \rceil$.[4] A strategy for i, $s_i \in S_i$, specifies an initial choice and a choice for each stage 1 history. We focus on pure strategies throughout. Finally, let $S = \times_{i \in N} S_i$ and $s \in S$.

The following observation highlights important properties of SPE of Γ_A.

[4] $\lceil \frac{\gamma}{\beta} \rceil$ is the lowest integer strictly greater than $\frac{\gamma}{\beta}$.

Observation 1 (Agreements game SPE)

(a) *There does not exist a full investment SPE in* Γ_A.

(b) *There exist zero investment SPE in* Γ_A*with* $[0, \lceil \frac{\gamma}{\beta} \rceil - 1)$ *signatories.*

(c) *There exist positive investment SPE in* Γ_A *with* $\lceil \frac{\gamma}{\beta} \rceil$ *signatories.*

Our proofs are rather tedious. We provide them in the appendix. In the main text we highlight key intuitions. In the case of Observation 1, to understand why SPE exhibit less than full investment, solve backwards. Non-signatories never invest as they bear their full investment cost. A signatory only invests if his costs are shared with sufficiently many co-signatories, namely if $m(a^1) \geq \lceil \frac{\gamma}{\beta} \rceil$. Given this, consider the sign-up stage. There exist SPE with very few signatories, (b), as stage 1 deviation does not induce investment. There also exist SPE with $\lceil \frac{\gamma}{\beta} \rceil$ signatories, (c), as each signatory is pivotal in inducing all signatories to invest. If there are greater than $\lceil \frac{\gamma}{\beta} \rceil$ signatories, a signatory can increase his payoff by deviating to not signing as other signatories continue to invest, hence full investment is not SPE, (a).

Observation 1 echoes the results of Varian (1994) and Barrett (1994).[5]

2.3 Modelling reciprocity

We now incorporate reciprocity following D&K.[6] Their approach uses "kindness functions". To determine whether i is kind to j one needs a reference point, the equitable payoff. In defining this reference point, D&K argue that only the set of efficient strategies, E_i, those not involving "wasteful" play, are relevant.[7] For Γ_P, $E_i = S_i$; for Γ_A, $s_i \notin E_i$ iff s_i prescribes not investing after a history of i signing and $m(a^1) \geq \lceil \frac{\gamma}{\beta} \rceil$.

Let $b_{ik} \in S_k$ denote i's (point) belief of k's strategy, then given $(b_{ik})_{k \neq i}$ the *equitable payoff* for j is the average of the most i believes he can "give"

[5]Both authors suggest mechanisms that give some SPE with partial investment and some with zero investment.

[6]Our presentation is tailored to Γ_P and Γ_A; see D&K for general games.

[7]A strategy is efficient if there does not exist another strategy which for all histories and others' strategies gives no player a lower payoff, and for some history and others' strategies gives at least one player a strictly higher payoff. See D&K pp. 275-6 for more on motivation, and the precise definition for general games.

to j given his strategy set and the least he believes he can "give" given his efficient strategies:

$$\pi_j^{e_i}\left((b_{ik})_{k\neq i}\right)=\frac{1}{2}\left[\max\{\pi_j\left(s_i,(b_{ik})_{k\neq i}\right)|s_i\in S_i\}+\min\{\pi_j\left(s_i,(b_{ik})_{k\neq i}\right)|s_i\in E_i\}\right].$$

To define i's kindness to j, let $s_i(h)$ be i's (updated) strategy which is identical to s_i except that a_i^1 must be consistent with reaching h. Let $b_{ik}(h)$ be i's (updated) belief of k's strategy. Given $s_i(h)$ and $(b_{ik}(h))_{k\neq i}$, i's *kindness* to j at h is

$$\kappa_{ij}\left(s_i(h),(b_{ik}(h))_{k\neq i}\right)=\pi_j\left(s_i(h),(b_{ik}(h))_{k\neq i}\right)-\pi_j^{e_i}\left((b_{ik}(h))_{k\neq i}\right).$$

The material payoff i believes j receives (first term on RHS) is compared to the equitable payoff (second term on RHS). If $\kappa_{ij}(.)>0$, i is kind to j. If $\kappa_{ij}(.)<0$, i is unkind to j. If $\kappa_{ij}(.)=0$, i has zero kindness toward j.

To capture reciprocity incentives D&K need a function reflecting how kind i perceives j as being. Let $c_{ijk}(h)$ denote i's updated (point) belief about j's (point) belief about k's strategy. Given $b_{ij}(h)$ and $(c_{ijk}(h))_{k\neq j}$, i's *perceived kindness* of j towards i at history h is

$$\lambda_{iji}\left(b_{ij}(h),(c_{ijk}(h))_{k\neq j}\right)=\pi_i\left(b_{ij}(h),(c_{ijk}(h))_{k\neq j}\right)-\pi_j^{e_j}\left((c_{ijk}(h))_{k\neq j}\right).$$

The material payoff that i believes j believes i receives (first term on RHS) is compared to the equitable payoff (second term on RHS). If $\lambda_{iji}(.)>0$ then i perceives j as kind to him, etc.

Player i 's utility at history h sums material and reciprocity payoffs

$$\begin{aligned}&U_i\left(s_i(h),\left(b_{ik}(h),(c_{ikl}(h))_{l\neq k}\right)_{k\neq i}\right)\\&=\underbrace{\pi_i\left(s_i(h),(b_{ik}(h))_{k\neq i}\right)}_{\text{material payoff}}\\&\underbrace{+Y\sum_{j\in N\backslash\{i\}}\left(\kappa_{ij}\left(s_i(h),(b_{ik}(h))_{k\neq i}\right)\cdot\lambda_{iji}\left(b_{ij}(h),(c_{ijl}(h))_{l\neq j}\right)\right)}_{\text{reciprocity payoff}},\end{aligned}\tag{1}$$

where $Y\geq 0$ is the sensitivity to reciprocity payoffs. If $Y>0$ then i's preference for reciprocation toward j is captured by i's utility increasing when $\kappa_{ij}(.)$ and $\lambda_{iji}(.)$ are non-zero with matching signs, reflecting mutual kindness or unkindness.

Let $\widehat{\Gamma}_P$ and $\widehat{\Gamma}_A$ denote the public good game and agreements game where players' utilities are (1).[8] To define an appropriate solution concept, let $s_i'(h)$ be the strategy identical to $s_i(h)$ at all histories except at h, where it differs (given binary action choices there is only one such strategy).

Definition 1 *$s \in S$ is a sequential reciprocity equilibrium (SRE) if for all i and at each history h it holds that*

$$\text{(i) } U_i\left(s_i(h), \left(b_{ik}(h), (c_{ikl}(h))_{l\neq k}\right)_{k\neq i}\right) \geq U_i\left(s_i'(h), \left(b_{ik}(h), (c_{ikl}(h))_{l\neq k}\right)_{k\neq i}\right),$$

(ii) $c_{ijk} = b_{jk} = s_k$ for all $j \neq i$ and $k \neq j$.

Condition (i) implies i best-responds at each history given his beliefs. Condition (ii) requires beliefs to be correct. If $Y = 0$ then Definition 1 describes a SPE (+ correct beliefs) in a game where utility equals material payoffs.

2.4 Reciprocity in the public good game

Full investment was impossible in Γ_P (2.1). Can it be a SRE in $\widehat{\Gamma}_P$? If so, given equilibrium beliefs, for all $i, j \in N$, $\kappa_{ij}(1,.) = \lambda_{iji}(1,.) = \frac{\beta}{2}$ and $\kappa_{ij}(0,.) = -\frac{\beta}{2}$. Therefore i would not deviate to not investing if $n\beta - \gamma + (n-1)Y(\frac{\beta}{2})^2 \geq (n-1)\beta + (n-1)Y(-\frac{\beta}{2})(\frac{\beta}{2})$. Put differently, full investment is a SRE of $\widehat{\Gamma}_P$ if $Y \geq Y^*$, where

$$Y^* = \frac{2(\gamma - \beta)}{\beta^2(n-1)}.$$

Intuitively, a sufficiently high reciprocity sensitivity ensures that the reciprocity cost of deviating to not investing (being unkind to co-players who are kind to you) outweighs the material gain. Our main result, in the next section, concerns whether the agreements mechanism ($\widehat{\Gamma}_A$) can deliver full investment where it is impossible in $\widehat{\Gamma}_P$, i.e. $Y < Y^*$.

[8] $\widehat{\Gamma}_P$ and $\widehat{\Gamma}_A$ are psychological games (Geanakoplos et al. 1989, Battigalli and Dufwenberg 2009).

3 Main result

Can the agreements game deliver full investment if this is impossible without the agreements mechanism ($Y < Y^*$) and the stakes are high (large β)? Indeed it can, but not how one might expect. If all players sign, full investment remains out of reach; however, if *no-one* signs, full investment *is* possible! We first state the result formally (3.1), then provide intuition via examples (3.2-3).

3.1 Formal statement

Theorem 1 *For all* $\gamma > 0$ *and* $n \geq 4$*, there exists* $\beta' \in (0, \gamma)$ *and* $Y' \in (0, Y^*)$ *such that,*

(a) *if* $\beta \geq \beta'$ *and* $Y < Y^*$*, there does not exist a full investment SRE in* $\widehat{\Gamma}_A$ *with* n *signatories,*

(b) *if* $\beta \geq \beta'$ *and* $Y \in [Y', Y^*)$*, there exists a full investment SRE in* $\widehat{\Gamma}_A$ *with* 0 *signatories. The SRE is described by* 0 *signing, then non-signatories investing iff there are* 0 *signatories and signatories investing if there are at least* $\lceil \frac{\gamma}{\beta} \rceil$ *signatories.*

The theorem suggests that for high return public goods where full investment is impossible, the introduction of an agreements mechanism makes full investment possible. Encouraging all players to actually sign however is counterproductive. Instead, it is potentially a good thing if all parties walk away from the negotiating table.

The strategies that support the full investment result – part (b) of the theorem – may seem to require a deep understanding of the game by the players. If a player deviates in stage 1 by signing, then no players invest in stage 2. So a deviation in stage 1 would result in a large decrease in material payoff for all players, and (as we show below) this is central for incentive-compatibility. Of course, this requires that all players understand the full consequences of varying degrees of take-up of the agreement, which may seem a tall order. However, on one interpretation, the equilibrium may reflect an *informal agreement*, where coordination is achieved through discussions. We return to and elaborate on that perspective later on (Section 5, Insight #2).

3.2 Highlighting the intuition: part (b)

We begin with part (b) of Theorem 1 as it is more straightforward. Example 1 takes the stated profile, where no-one signs and all invest on path, and shows that it is a SRE for an interval of Y less than Y^*.

Example 1 (No-one signs): Take $\widehat{\Gamma}_A$ and let $n = 4$, $\gamma = 10$, $\beta = 9$ and $Y \in [\frac{1}{351}, Y^*)$, where $Y^* = \frac{2}{243}$. Consider the profile where,

- no-one signs,
- non-signatories invest iff there are 0 signatories,
- signatories invest iff there are at least 2 signatories.

Reason as follows to see that the profile is a SRE. Following a history of 0 signatories, i does not deviate to not investing if

$$\begin{aligned} &4 \cdot 9 - 10 + 3Y \cdot \left(26 - \tfrac{1}{2}(26+0)\right) \cdot \underbrace{\left(26 - \tfrac{1}{2}(26+0)\right)}_{\lambda_{iji}(.)} \\ &\geq 3 \cdot 9 + 3Y \cdot \left(17 - \tfrac{1}{2}(26+0)\right) \cdot \underbrace{\left(26 - \frac{1}{2}(26+0)\right)}_{\lambda_{iji}(.)}. \end{aligned} \tag{2}$$

Solving gives $Y \geq \frac{1}{351}$, note that $\frac{1}{351} < Y^* = \frac{2}{243}$.

Following a history of 1 signatory, i does not deviate to investing for any Y. Following a history of 2 signatories, signatory i does not deviate to not investing if $Y \leq \frac{4}{157}$, note that $Y^* = \frac{2}{243} < \frac{4}{157}$. Following histories of 2 or 3 signatories, non-signatory i does not deviate to investing if $Y \leq Y^*$. Following a history of 3 or 4 signatories, signatory i does not deviate to not investing for any Y. Player i does not deviate to signing in the sign-up stage for any Y. The profile is thus a SRE for $Y \in [\frac{1}{351}, Y^*)$. ▲

Kindness amplification

Why do players invest after no-one signs (Example 1), when they would not have invested if there had been no option to sign ($\widehat{\Gamma}_P$ with $Y < Y^*$)? The explanation centers on what we call *kindness amplification*.

To see this, contrast the 0 signatory subgame in Example 1 with the full investment profile in $\widehat{\Gamma}_P$. In both cases, i has the same material incentive to deviate to not investing: a material payoff increase of $3\cdot9-(4\cdot9-10)=1$. His reciprocity incentive however differs as kindness is higher (or amplified) in Example 1 (e.g. $\lambda_{iji}(.)=\frac{9}{2}$ in $\widehat{\Gamma}_P$, see section 2.4, but $\lambda_{iji}(.)=13$ in Example 1, see (2)). The amplified kindness implies that for a given Y, the decrease in i's reciprocity payoff from deviating to not investing is larger in Example 1 ($3Y\cdot13\cdot13-3Y\cdot4\cdot13=351Y$) than in $\widehat{\Gamma}_P$ ($3Y(\frac{9}{2})(\frac{9}{2})-3Y(-\frac{9}{2})(\frac{9}{2})=\frac{243}{2}Y$). A lower Y is thus sufficient to prevent i deviating in Example 1 than in $\widehat{\Gamma}_P$, hence our result.

The fundamental reason as to why the agreements mechanism can amplify kindness is that it can increase a player's ability to influence others' behaviour. For instance, if j signs in Example 1, *no-one* invests, implying i's material payoff falls by $4\cdot9-10=26$. If j deviates to not investing in $\widehat{\Gamma}_P$, others' actions are unchanged so i 's material payoff falls by only $9<26$. Thus i perceives j as kinder in Example 1 than in $\widehat{\Gamma}_P$.

3.3 Highlighting the intuition: part (a)

We now illustrate part (a) of Theorem 1. Example 2 has three parts, it takes the parameters of Example 1 and demonstrates that for $Y<Y^*$, no profile where all players sign and invest on path is a SRE. The key intuition is that kindness cannot be amplified as effectively as when there are zero signatories on path (3.2). Part (i) demonstrates that a particular off-path investment-stage behaviour implies sign-up stage deviation. Parts (ii) and (iii) show that while alternative off-path investment-stage behaviours may avoid sign-up stage deviation, they necessarily involve investment-stage deviation.

Example 2(i) (Sign-up stage deviation): Take $\widehat{\Gamma}_A$ and let $n=4$, $\gamma=10$, $\beta=9$ and $Y<Y^*=\frac{2}{243}$. Consider any profile where

- all players sign & invest on path,
- only signatories invest if there are 3 signatories.

Player i's sign-up stage incentives are identical to those for the full investment profile in $\widehat{\Gamma}_P$, i thus deviates to not signing as $Y<Y^*$. ▲

It follows from Example 2(i) that all remaining candidate SRE (where all players sign and invest on path) must involve: (a) a signatory who does not invest when there are 3 signatories, or, (b) a non-signatory who invests when there are 3 signatories. Examples 2(ii) and 2(iii) rule out candidate SRE with properties (a) and (b) respectively by demonstrating that players deviate in the investment stage.

Example 2(ii) (Signatory not investing with 3 signatories): Take $\widehat{\Gamma}_A$ and let $n = 4$, $\gamma = 10$, $\beta = 9$ and $Y < Y^* = \frac{2}{243}$. Consider a profile where

- all players sign & invest on path,
- at history h where there are 3 signatories, some signatory i does not invest.

By deviating to investing at h, signatory i can increase his material payoff by $9 - \frac{10}{3} = \frac{17}{3}$. Can i's reciprocity incentives prevent this deviation? Since i not investing is less kind than investing, to maximise the reciprocity cost of i's deviation examine the profile where

- all players sign & invest on path,
- no-one invests if there are 3 signatories,
- all players invest if there are 2 signatories.

Player i's reduction in reciprocity payoff from deviation at h is then $\frac{1144}{3}Y < \frac{2288}{729} < \frac{17}{3}$ (first inequality from $Y < Y^* = \frac{2}{243}$), thus i deviates. ▲

Weak kindness amplification

Why can kindness be sufficiently amplified to prevent deviation in Example 1 but not in Example 2(ii)? The answer concerns the size of the material loss that the reciprocity payoff need compensate in the relevant subgames. In Example 1, following a history of zero signatories, the material loss of investing was 1. In Example 2(ii), following a history of 3 signatories, the material loss to a signatory of not investing was $\frac{17}{3} > 1$. More generally, for high β, the size of the material loss from not deviating in the relevant subgame is lower for the case when there are zero signatories than when there are many. Lower kindness is thus sufficient for non-deviation when there are few signatories than when there are many.

To complete Example 2, part (iii) examines the remaining candidate SRE (where all sign and invest on path) which involve all players investing when there are 3 signatories (by Examples 2(i)-(ii)). It shows that for a non-signatory to invest when there are 3 signatories, a signatory must not invest when there are 2 signatories. However, since this signatory has an incentive to deviate to investing, these profiles are not SRE either.

Example 2(iii) (Non-signatory investing with 3 signatories): Take $\widehat{\Gamma}_A$ and let $n = 4$, $\gamma = 10$, $\beta = 9$ and $Y < Y^* = \frac{2}{243}$. Consider a profile where

- all players sign & invest on path,
- all players invest at h where all but i have signed

By deviating to not investing at h, non-signatory i increases his material payoff by $10 - 9 = 1$. Whether i's reciprocity incentives prevent deviation depends on behaviour following histories with 2 signatories. If all invest when there are 2 signatories, i's incentives are identical to the full investment profile in $\widehat{\Gamma}_P$, thus i deviates at h ($Y < Y^*$). If only the signatories invest when there are 2 signatories, then i deviates at h if $Y < \frac{1}{108}$, which is true ($Y < Y^* = \frac{2}{243} < \frac{1}{108}$). For i not to deviate at h, it must be that at least 3 players do not invest when there are 2 signatories.

Suppose then that some signatory j does not invest when there are 2 signatories. Signatory j has a material incentive to deviate to investing. To maximise the reciprocity cost of this deviation, suppose no-one invests if there are 2 signatories and all invest if there is 1 signatory. Signatory j deviates to investing when there are 2 signatories if $Y < \frac{2}{143}$, which is true ($Y < Y^* = \frac{2}{243} < \frac{2}{143}$). ▲

For the non-signatory to invest when there are 3 signatories his perceptions of others' kindness must be sufficiently amplified, this requires signatories to not invest when there are 2 signatories. However, reciprocity incentives are not sufficiently large to prevent a signatory deviating to investing for reasons analogous to those discussed following Example 2(ii). Thus there does not exist a full investment SRE where all sign on path.

To summarise, the fundamental intuition behind part (a) of Theorem 1 is the difficulty of amplifying kindness when many players sign. In order to prevent sign-up stage deviation, kindness must be amplified, this requires low investment in subgames where there are $n-1$ or $n-2$ (3 or 2 in Example 2) signatories. However, signatories have relatively large material incentives to invest in such subgames (high β) and reciprocity payoffs are relatively low ($Y < Y^*$), hence the impossibility.

4 Further results

This section addresses three further questions. Does the impossibility of full investment with n signatories arise in 2- and 3-player games? Can the agreements mechanism give full investment when it is possible without a mechanism ($Y \geq Y^*$)? When reciprocity is low ($Y \in (0, Y^*)$) can the mechanism deliver at least the investment levels possible under material preferences? We show that the answers are no, yes and it depends.

Three- and two-player games

In contrast to Theorem 1(a), for 3-player games, the agreements mechanism can give full investment with n signatories for an interval of Y less than Y^*.[9]

Proposition 1 (3-players) *For $n = 3$, all β and γ, there exists $Y'' \in (0, Y^*)$ such that,*

(a) *if $Y \in [Y'', Y^*)$ there exists a full investment SRE in $\widehat{\Gamma}_A$ with 3 signatories. The SRE is described by 3 signing, then i does not invest iff there is only 1 signatory.*

(b) *if $Y \in [Y'', Y^*)$ there exists a full investment SRE in $\widehat{\Gamma}_A$ with 0 signatories. The SRE is described by 0 signing, then i does not invest iff there is only 1 signatory.*

The impossibility identified in Section 3 does not arise here as kindness can be amplified more easily when $n = 3$. From Section 3, recall that a candidate SRE where n sign and invest on path, required low investment in subgames where there were $n-1$ or $n-2$ signatories (see Example 2(ii)-(iii)). Large material incentives to invest in such subgames prevented low investment for $n > 3$. By contrast, for $n = 3$, we get $n - 2 = 1$ signatory and players have no material incentive to invest if there is only one signatory. Kindness can thus be sufficiently amplified and all players signing and investing on path is a SRE (Proposition 1(a)).

Now consider 2-player games. For such games, drop the assumptions that $\frac{\gamma}{\beta}$ is non-integer and $n > \left\lceil \frac{\gamma}{\beta} \right\rceil$ as they would contradict that $n\beta > \gamma > \beta > 0$.[10]

[9] From now on, allow Γ_A and Γ_P to be 2- or 3-player games.

[10] Note that there exists a full investment SPE in the 2-player Γ_A where both sign and invest on path.

Proposition 2 (2-players) *For $n = 2$, all γ and $Y \in (0, Y^*)$,*

(a) *if $\beta \geq \frac{2}{3}\gamma$, then there exists a full investment SRE in $\widehat{\Gamma}_A$ with 2 signatories. The SRE is described by 2 signing, then i invests iff there are 2 signatories.*

(b) *for all β, there does not exist a full investment SRE in $\widehat{\Gamma}_A$ with 0 signatories.*

In the 2-player agreements game, there exists a full investment SPE with 2 signatories. Kindness amplification is not needed to compensate players for a material loss, thus impossibility does not occur (part (a)). By contrast, full investment with zero signatories does require kindness amplification. As part (b) suggests, kindness cannot be sufficiently amplified in the 2-player game, this is because there is no other player, k, such that i can influence k's strategy and thereby amplify i's influence over j's material payoff.

High reciprocity ($Y \geq Y^*$)
Our main result examined cases where full investment was impossible with no mechanism ($Y \in (0, Y^*)$). We now consider whether the mechanism precludes full investment when it was possible with no mechanism ($Y \geq Y^*$).

Proposition 3 (High reciprocity) *For all $n \geq 2$, γ, β and $Y \geq Y^*$ there exists a full investment SRE in $\widehat{\Gamma}_A$. The SRE is described by n signing, then i does not invest iff there are $\lceil \frac{\gamma}{\beta} \rceil - q$ signatories where $q > 0$ and odd.*

If full investment were possible without the mechanism, it remains possible with it. As the reciprocity sensitivity is high ($Y > Y^*$), when there are many signatories ($\geq \lceil \frac{\gamma}{\beta} \rceil$), reciprocity payoffs compensate material costs of investment with no need for kindness amplification. When there are few signatories ($\leq \lceil \frac{\gamma}{\beta} \rceil$), the alternating between all investing and zero investing ensures kindness is sufficiently amplified and that there are no deviations.

Low reciprocity ($Y \in (0, Y^*)$)
Observation 1(a) stated that full investment was impossible with material preferences. Our results have thus far considered whether it is attainable with reciprocity. We found that full investment is indeed attainable if the reciprocity sensitivity is high enough (e.g. Theorem 1, Proposition 3). What about when the reciprocity sensitivity is not sufficiently high? We know

from Observation 1(c) that even with material preferences (i.e. $Y = 0$), equilibria exist with $\lceil \frac{\gamma}{\beta} \rceil$ players investing. Do equilibria exist with at least $\lceil \frac{\gamma}{\beta} \rceil$ players investing for small, non-zero reciprocity sensitivities? The answer is often yes, however possibilities are complicated. We do not give a complete characterisation, but offer a sufficient condition for games with many players.

Proposition 4 (Low reciprocity) *For all $n \geq 7$, γ and $Y \in (0, Y^*)$, there exists $\beta'' \in (0, \gamma)$ and $m'(Y) \in \left[\lceil \frac{\gamma}{\beta} \rceil, n\right]$ such that*

- ***(a)*** *if $\beta \geq \beta''$ there exists a SRE where $m'(Y)$ players sign and invest on path. The SRE is described by $m'(Y)$ players signing, then i invests iff i signed and there are at least $m'(Y)$ signatories.*
- ***(b)*** *$m'(Y)$ is non-decreasing in Y.*

For $n \geq 7$, even when the reciprocity sensitivity is low ($Y \in (0, Y^*)$) at least the maximal SPE investment, $\lceil \frac{\gamma}{\beta} \rceil$, is attainable as a SRE. As Y increases, investment greater than $\lceil \frac{\gamma}{\beta} \rceil$ is possible in SRE, since reciprocity payoffs increase due to mutual kindness among signatories and kindness amplification.

5 Conclusion

We started this project because we felt that the economic situations described by the agreements game were plentiful and important, and that reciprocity motivation was realistic in those settings. If game form, utilities and solution concepts make sense, so should the results. We were astonished when we stumbled upon Theorem 1, which we found extremely surprising.

At first we were at loss for real-world examples that the result might shed light on. But then two insights, both concerning the nature of relevant observations to look for, dawned on us:

Insight #1: The dog that didn't bark
Relevant non-agreement examples may easily be missed because they are *counterfactual*. For example, consider a set of countries (e.g. China and Malaysia, or Norway, Iceland and the UK) separated by a sea with a limited fish stock. A tragedy-of-the-commons over-fishing scenario may seem ominous. What we call "investing" would be analogous to holding back on

sending out trawlers. A co-financing agreement would involve compensating neighbouring countries for their opportunity cost of catch they didn't get. Now, suppose such an agreement is *not* struck. Would you have noticed? The answer may be yes, if there were lengthy negotiations that eventually stranded, with upset politicians who cursed at each other during press conferences. However, a no-one-signs outcome may also involve no drama at all. It might be, that a co-financing agreement didn't happen simply because no one ever tried to move in that direction. For the analyst, it is, at first, easy to overlook and not ever notice such non-agreements. But in fact, once one starts thinking about them, one realises that deals-that-do-not-happen happen all the time.

The UN, for example, offers an institutional setting where many agreements could be struck, in principle, and, often, are not. The US and Russia have not signed a co-financing agreement to rebuild Syria, for instance, and their diplomats may have never raised the issue. It is easy to be critical of that example: Is the situation really like a public good game? Well, maybe. Is the outcome we see in Syria actually any good, as Theorem 1 would have it? That may seem a stretch, but consider that the situation could conceivably have been considerably worse (e.g. World War III). Wary of over-reach, we shall not over-state our case. And perhaps the fishing example is more compelling than the Syria example. But we propose that in either case the issue is thought-provoking and stimulating to consider.

Insight #2: Informal agreements

A weakness of Insight #1 is that non-deals that no-one ever moved towards may not have been perceived by the parties supposed to play the game. That would erode the applicability of our theory, which assumes that the parties recognise that they interact in a public good game form, where a co-financing agreement is a real possibility even when shunned. Our second insight avoids this weakness:

Namely, a no-one signs outcome may follow neither a stranded negotiation nor a no-one-did-anything scenario. Rather, the outcome may have been quite actively and harmoniously reached, with a co-financing agreement being explicitly rejected while some other informal and non-binding agreement (corresponding to full investment without co-financing) were reached. As a possible example, take the 2015 UN Climate Change Conference in Paris (aka COP21/CMP11). Many commentators have pointed out that the agreement is not binding, in the sense that there are no penalties for non-compliance (e.g. Clémençon (2016), Jacquet and Jamieson (2016)). Arguably, negotiators could have made a binding, or formal, commitment to co-finance each other's efforts to reduce greenhouse gas emissions, but instead they struck a non-binding, or informal, agreement to exert such effort anyway, without co-financing mandates (at least between industrial countries). This may be an outcome along the lines that our Theorem 1 points to. Agents do not sign a formal co-financing agreement, but rather reject the agreement and coordinate on full investment nonetheless. The "soft-touch" Paris agreement is then a much needed coordination device to ensure that investments in climate change are made. On this interpretation, Theorem 1 is less about non-deals, and more about allowing certain informal deals to gain traction as other formal deals are shunned. And institutions that allow formal agreements for co-financing to be signed, even when they are not, are critical for spurring informal agreements to invest in public goods.[11]

Time will tell whether the climate is in good hands following the Paris agreement. According to our model, under the interpretation given here, there are some intriguing reasons to be optimistic.

Insights #1 and #2 highlight a general lesson: Institutions that offer opportunities for binding agreements may be important even if such agreements are not struck. In our case, creating the possibility of a formal co-financing agreement may promote actual investment in public goods, and yet no co-financing agreement is signed.

[11]Environmental economist Michael Greenstone (2015; online NYT article) recently argued that international treaties (and in particular the Paris one) have to be voluntary and such that individual countries find compliance in their interest. The insights of our Theorem 1 may be seen to complement his view, endogenising the nature of said agreement and providing an account for the forces (*viz.* reciprocity) that ensure incentive compatibility.

Appendix

Proof of Observation 1 (Agreements game SPE)

Apply backward induction to identify the SPE of Γ_A. At $h = a^1$, non-signatories do not invest and signatory i invests iff $\beta \geq \frac{\gamma}{m(a^1)}$. Since $\frac{\gamma}{\beta}$ non-integer, write the condition as iff $m(a^1) \geq \lceil \frac{\gamma}{\beta} \rceil$. Given optimal behaviour at all a^1, consider the first-stage. First suppose there are less than $\lceil \frac{\gamma}{\beta} \rceil - 1$ signatories. Player i does not deviate in the first-stage as $\pi_i(.) = 0$ regardless. Thus this is a SPE. Second suppose there are $\lceil \frac{\gamma}{\beta} \rceil - 1$ signatories. Non-signatory i deviates to signing if $0 < \beta \lceil \frac{\gamma}{\beta} \rceil - \gamma$, which is true. Thus this is not a SPE. Third suppose there are $\lceil \frac{\gamma}{\beta} \rceil$ signatories. Signatory i does not deviate to not signing if $\lceil \frac{\gamma}{\beta} \rceil \beta - \gamma \geq 0$, which is true. Non-signatory i does not deviate to signing if $\lceil \frac{\gamma}{\beta} \rceil \beta \geq (\lceil \frac{\gamma}{\beta} \rceil + 1)\beta - \gamma$, which is also true. Thus this is a SPE. Fourth suppose there are more than $\lceil \frac{\gamma}{\beta} \rceil$ signatories. Signatory i deviates to not signing as other players invest regardless of his choice. Thus this is not a SPE. The four cases are exhaustive. ■

Proof of Theorem 1

(a) Take the set of strategy profiles that involve n players signing and investing on path. Partition this set into 3 subsets of profiles distinguished by behaviour following a history of a^1 such that $m(a^1) = n - 1$: *subset 1*, all invest; *subset 2*, all signatories invest only; and *subset 3*, all remaining profiles. We take each subset in turn and demonstrate that no profile in the subset is a SRE if β is sufficiently high and $Y < Y^*$.

Subset 1: Consider any candidate SRE profile s^* such that each $i \in N$ signs, then invests if a^1 is such that $m(a^1) \geq n - 1$. Reason as follows to show that there is no behaviour at histories such that $m(a^1) < n - 1$ that would imply s^* is a SRE.

Consider $h = a^1$ such that $m(a^1) = n - 1$, so all players invest. Non-signatory i has the same material incentive to deviate to not investing as in Γ_P. Given $Y < Y^*$, a necessary condition for i not to deviate is that $\lambda_{iji}(s_j^*, .) > \frac{\beta}{2}$ (recall $\lambda_{iji}(1, .) = \frac{\beta}{2}$ in $\widehat{\Gamma}_P$). The value of $\lambda_{iji}(s_j^*, .)$ at h, depends on the action choices s^* prescribes at history h' where all except i and j sign. If s^* were such that n invest or all except j invest at h', then $\lambda_{iji}(s_j^*, .) = n\beta - \gamma - \frac{1}{2}(n\beta - \gamma + (n-1)\beta - \gamma) = \frac{\beta}{2}$ at h, thus i would deviate at h. If s^* were such that all except i invest or all except i and j invest at h', then $\lambda_{iji}\left(s_j^*, .\right) = n\beta - \gamma - \frac{1}{2}\left((n-1)\beta + (n-2)\beta\right) = \frac{3}{2}\beta - \gamma < \frac{\beta}{2}$, thus i would deviate at h. Therefore a necessary condition for i to not deviate at h, is that s^* must be such that some signatory l does not invest at h'.

Consider h' and suppose s^* prescribes signatory l does not invest. Signatory l has a material incentive to deviate to invest at h'. We now show that for β sufficiently high and all $Y < Y^*$, l's reciprocity incentives are insufficient to prevent deviation at h'. The change in signatory l's reciprocity payoff from playing s_l^* rather than $s_l'(h', s_l^*)$ equals $Y\Psi$, where

$$\Psi := \sum_{k \in N \setminus \{l\}} \left(\kappa_{lk}\left(s_l^*, .\right) - \kappa_{lk}\left(s_l'\left(h', s_l^*\right), .\right)\right) \lambda_{lkl}\left(s_k^*, .\right).$$

If $\Psi \leq 0$, then l deviates at h'. Suppose $\Psi > 0$. Signatory l does not deviate at h' if

$$Y \geq \frac{(n-2)\beta - \gamma}{(n-2)\Psi}. \tag{3}$$

Let $\widehat{Y}(\beta)$ denote the RHS of (3) as a function of β. For $Y < Y^*$ we require

$$\widehat{Y}(\beta) = \frac{(n-2)\beta - \gamma}{(n-2)\Psi} < \frac{2(\gamma - \beta)}{\beta^2(n-1)} = Y^*.$$

Now argue that for sufficiently high β either $Y < Y^*$ does not hold or l has an incentive to deviate at h'. Note that $\lim_{\beta \to \gamma} Y^* = 0$. Therefore for β in the neighbourhood of γ, $Y < Y^*$ requires $\lim_{\beta \to \gamma} \widehat{Y}(\beta) \leq 0$. Evaluating $\widehat{Y}(\gamma)$, note that the numerator of $\widehat{Y}(\gamma)$ is positive thus it must be that $\Psi \leq 0$. However if $\Psi < 0$ then l would deviate at h' for β slightly lower than γ as already argued, thus $\Psi = 0$ when $\beta = \gamma$. That we have supposed that $\Psi > 0$ for sufficiently high β and deduced $\Psi = 0$ at $\beta = \gamma$, implies that the denominator of $\widehat{Y}(\beta)$ approaches zero from above and hence the one-sided limit $\lim_{\beta \to \gamma^-} \widehat{Y}(\beta) = +\infty$ which is greater than $\lim_{\beta \to \gamma} Y^* = 0$, violating $Y < Y^*$. Thus for all $Y < Y^*$ and β sufficiently high, l would deviate to investing at h'.

Subset 2: Consider a candidate SRE profile s^* such that each $i \in N$ signs, then invests if a^1 is such that $m\left(a^1\right) = n$, and if a^1 is such that $m\left(a^1\right) = n - 1$, all except the non-signatory invest. At the initial node, i's incentives are identical to those in the full investment profile in $\widehat{\Gamma}_P$. Thus i deviates to not signing at the initial node for all $Y < Y^*$. Hence s^* is not a SRE.

Subset 3: Consider any remaining candidate SRE profile s^* such that each $i \in N$ signs, then invests if a^1 is such that $m(a^1) = n$, it must be that for history $h'' = a^1$ such that $m(a^1) = n - 1$, there exists some signatory r who does not invest. Reasoning analogous to that used to show signatory l deviates to investing at h' (end of subset 1) establishes that signatory r deviates to investing at h''. Hence s^* is not a SRE.

(b) Consider s^* such that each $i \in N$ does not sign, then invests if a^1 is such that $m(a^1) = 0$ or $a_i^1 = 1$ and $m(a^1) \geq 2$, and does not invest otherwise. We demonstrate that there exists $Y'' < Y^*$ such that no player deviates at any h if $Y \in [Y'', Y^*)$ and β is sufficiently large.

First consider $h = a^1$ such that $m(a^1) = 0$. Player i has a material incentive to deviate to not investing and a reciprocity incentive to not do so. His increase in reciprocity payoff from playing s_i^* instead of $s_i'(h, s_i^*)$ is

$$\begin{aligned}&\left(n - m(a^1) - 1\right) \cdot Y \cdot \left(\kappa_{ij}(s_i^*, .) - \kappa_{ij}(s_i'(h, s_i^*), .)\right) \cdot \lambda_{iji}\left(s_j^*, .\right) \\ &+ m(a^1) \cdot Y \cdot \left(\kappa_{il}(s_i^*, .) - \kappa_{il}(s_i'(h, s_i^*), .)\right) \cdot \lambda_{ili}(s_l^*, .),\end{aligned} \tag{4}$$

where j is a non-signatory and l a signatory. Since $\kappa_{ij}(s_i^*, .) = \lambda_{iji}\left(s_j^*, .\right) = \frac{1}{2}(n\beta - \gamma)$ and $\kappa_{ij}(s_i'(h, s_i^*)) = \frac{1}{2}(n\beta - \gamma) - \beta$, this increase in reciprocity payoff is larger than the reduction in material payoff if $Y \geq \Phi$, where

$$\Phi := \frac{2(\gamma - \beta)}{(n-1)\beta(n\beta - \gamma)}.$$

Note that if $\beta > \frac{\gamma}{n-1}$, then $\Phi < Y^*$. Thus i does not deviate for $Y \in [\Phi, Y^*)$.

Now consider $h = a^1$ such that $m(a^1) > 0$. No player has a material incentive to deviate at a^1 since for $\beta > \frac{\gamma}{2}$, $\lceil \frac{\gamma}{\beta} \rceil = 2$, thus action choices following all a^1 are identical to SPE profiles. We will demonstrate for sufficiently high β and Y (but less than Y^*), any reciprocity incentive to deviate is less than the material incentive to not do so.

At $h = a^1$ such that $m(a^1) = 1$, player i has no reciprocity incentive to deviate to investing, thus does not deviate. At $h = a^1$ such that $m(a^1) = 2$, the change in signatory i's reciprocity payoff from playing s_i^* rather than $s_i'(h, s_i^*)$ is

$$\begin{aligned}&\left(n - m(a^1)\right) \cdot Y \cdot \left(\kappa_{ij}(s_i^*, .) - \kappa_{ij}(s_i'(h, s_i^*), .)\right) \cdot \lambda_{iji}\left(s_j^*, .\right) \\ &+ \left(m(a^1) - 1\right) \cdot Y \cdot \left(\kappa_{il}(s_i^*, .) - \kappa_{il}(s_i'(h, s_i^*), .)\right) \cdot \lambda_{ili}(s_l^*, .),\end{aligned} \tag{5}$$

where j is a non-signatory and l a signatory. For non-signatory j, $\kappa_{ij}(s_i^*, .) = \beta$, $\kappa_{ij}(s_i'(h, s_i^*), .) = 0$ and $\lambda_{iji}(s_j^*, .) = -\frac{\beta}{2}$. For signatory l, $\kappa_{il}(s_i^*, .) = \frac{3}{2}\beta - \gamma$, $\kappa_{il}(s_i'(h, s_i^*), .) = \frac{1}{2}(\beta - \gamma)$ and $\lambda_{ili}(s_l^*, .) = \frac{3}{2}\beta - \gamma$. Substituting into (5) gives $Y\left(\left(\beta - \frac{\gamma}{2}\right)\left(\frac{3}{2}\beta - \gamma\right) - (n-2)\frac{\beta^2}{2}\right)$, which is negative for sufficiently large β. Signatory i's reciprocity incentive to deviate to not investing at h is no larger than the material incentive to not do so if

$$Y \leq \left[\frac{\frac{\gamma}{2} - \beta}{\left(\beta - \frac{\gamma}{2}\right)\left(\frac{3}{2}\beta - \gamma\right) - (n-2)\frac{\beta^2}{2}}\right].$$

For β sufficiently large, there exists Y satisfying the inequality and that $Y \geq \Phi$ (as $\lim_{\beta \to \gamma}[.] > 0$ and $\lim_{\beta \to \gamma}\Phi = 0$).

Using (4), non-signatory i's change in reciprocity payoff from playing s_i^* rather than $s_i'(h, s_i^*)$ is $\frac{1}{2}Y\beta^2(n-7)$. This is non-negative for $n \geq 7$, thus i does not deviate at h. For $n \in [4, 6]$, i does not deviate at h if $Y \leq \frac{2(\gamma - \beta)}{\beta^2(7-n)}$. There exists Y satisfying the inequality and that $Y \geq \Phi$, since the RHS of this inequality is greater than Φ if $(n^2 - 7)\beta + (1-n)\gamma > 0$, which holds given $n \geq 4$.

Consider $h = a^1$ such that $m(a^1) \in [3, n-1]$. Using (5), signatory i's change in reciprocity payoff from playing s_i^* rather than $s_i'(h, s_i^*)$ is $Y\Xi$ where

$$\Xi := \left(m\left(a^1\right) - 1\right)\left(\beta - \frac{\gamma}{m\left(a^1\right)}\right)\frac{\beta}{2} - \left(n - m\left(a^1\right)\right)\frac{\beta^2}{2}.$$

If $\Xi \geq 0$, then i has no reciprocity incentive to deviate at h. Suppose $\Xi < 0$, then i does not deviate if $Y \leq \left[\frac{\gamma - \beta m(a^1)}{\Xi m(a^1)}\right]$. For β sufficiently large, there exists Y satisfying the inequality and that $Y \geq \Phi$ (as $\lim_{\beta \to \gamma}[.] > 0 > \lim_{\beta \to \gamma}\Phi = 0$).

Non-signatory i's change in reciprocity payoff from playing s_i^* rather than $s_i'(h, s_i^*)$ is $\frac{1}{2}Y\beta^2(n - 2m(a^1) - 1)$ (use (4)). If $n - 2m(a^1) - 1 \geq 0$, then $\frac{1}{2}Y\beta^2(n - 2m(a^1) - 1) \geq 0$, thus i has no reciprocity incentive to deviate at h. If $n - 2m(a^1) - 1 < 0$, then i does not deviate at h if

$$Y \leq \frac{2(\gamma - \beta)}{\beta^2(2m(a^1) + 1 - n)}.$$

The RHS is strictly greater than Φ if $(n^2 - 2m(a^1) - 1)\beta + (1-n)\gamma > 0$ which is true as β tends to γ.

Finally, at $h = a^1$ such that $m(a^1) = n$ and the initial node, player i has neither material nor reciprocity incentives to deviate. ■

Proof of Proposition 1 (3-players)

(a) Let $n = 3$. Consider s^* such that each $i \in N$ signs, then does not invest if a^1 is such that $m(a^1) = 1$ and does invest otherwise. Reason as follows to verify the profile is SRE for an interval of Y less than Y^*. Consider $h = a^1$ such that $m(a^1) = 3$, so all invest. Signatory i has neither material nor reciprocity incentive to deviate to not investing. Now consider $h = a^1$ such that $m(a^1) = 2$, so all invest. Signatory i has neither a material nor a reciprocity incentive to deviate to not investing at h. Non-signatory i has a material incentive to deviate to not investing at h and a reciprocity incentive to not do so. Reason as follows to identify Y such that i does not deviate. Using (4), non-signatory i's increase in reciprocity payoff from playing s_i^* rather than $s_i'(h, s_i^*)$ is no less than his reduction in material payoff if $Y \geq Y''(\beta, \gamma)$, where

$$Y''(\beta, \gamma) := \frac{\gamma - \beta}{\beta(3\beta - \gamma)}.$$

Note that $Y''(\beta, \gamma) \geq Y^*$ iff $\frac{\gamma}{\beta} + 1 > 3$, however given $\lceil \frac{\gamma}{\beta} \rceil < n = 3$, then $Y''(\beta, \gamma) < Y^*$. Thus i does not deviate at h if $Y \in (Y'', Y^*)$. Now consider $h = a^1$ such that $m(a^1) = 1$, so zero invest. Player i has neither a material nor a reciprocity incentive to deviate to investing. Then consider $h = a^1$ such that $m(a^1) = 0$, so all players invest. Non-signatory i, faces identical incentives as a non-signatory at a history with 2 signatories, i does not deviate at h if $Y \in (Y'', Y^*)$. Finally, at the initial node, i has neither reciprocity nor material incentives to deviate.

(b) Let $n = 3$. Consider s^* such that each $i \in N$ does not sign, then does not invest if a^1 is such that $m(a^1) = 1$ and does invest otherwise. Reason as follows to verify the profile is SRE for an interval of Y less than Y^*. Stage 2 behaviour is optimal (part (a) of this proof). At the initial node, i has neither reciprocity nor material incentives to deviate. ■

Proof of Proposition 2 (2-players)

(a) Let $n = 2$. Consider s^* such that each $i \in N$ signs, then invests iff a^1 is such that $m(a^1) = 1$. Note that i has no material incentive to deviate at any history. Furthermore if $\beta \geq \frac{2}{3}\gamma$, i has no reciprocity incentive to deviate either.

(b) Let $n = 2$. Consider any s^* such that each $i \in N$ does not sign, then invests on path. Consider $h = a^1$ such that $m(a^1) = 0$, so all invest. Non-signatory i has a material incentive to deviate to not investing. Non-signatory i's increase in reciprocity payoff from playing s_i^* rather than $s_i'(h, s_i^*)$ is (4). We now demonstrate that this increase is reciprocity payoff is not sufficient to prevent deviation for $Y < Y^*$. Consider the following 4 exhaustive cases.

Case (i): s^* is such that j does not invest if only i signs and i does not invest if only j signs. Note that (4) is no less than the reduction in i's material payoff from playing s_i^* rather than $s_i'(h, s_i^*)$ iff $Y > \frac{2(\gamma-\beta)}{\beta(2\beta-\gamma)}$. However the RHS is less than Y^* iff $\beta > \gamma$, which is false.

Case (ii): s^* is such that j does not invest if only i signs and i does invest if only j signs. Note that (4) is no less than the reduction in i's material payoff from playing s_i^* rather than $s_i'(h, s_i^*)$ iff $Y \geq Y^*$, which is false.

Case (iii): s^* is such that j does invest if only i signs and i does not invest if only j signs. Note that (4) is no less than the reduction in i's material payoff from playing s_i^* rather than $s_i'(h, s_i^*)$ iff $Y > \frac{2(\gamma-\beta)}{\beta(2\beta-\gamma)}$. However the RHS is less than Y^* iff $\beta > \gamma$, which is false.

Case (iv): s^* is such that j does invest if only i signs, i does invest if only j signs. Note that (4) is no less than the reduction in i's material payoff from playing s_i^* rather than $s_i'(h, s_i^*)$ iff $Y \geq Y^*$, which is false.

Thus i deviates at h. ■

Proof of Proposition 3 (High reciprocity)

We demonstrate that a particular profile implying full investment, s^*, is a SRE of $\widehat{\Gamma}_A$ for all $Y \geq Y^*$. Consider s^* such that each $i \in N$ signs, then does not invest if a^1 is such that $m(a^1) = \lceil \frac{\gamma}{\beta} \rceil - q$ where $q > 0$ and odd, and does invest otherwise. Reason as follows to confirm that for all $Y \geq Y^*$, players have no incentive to deviate at any history.

Consider $h = a^1$ such that $m(a^1) \in (\lceil \frac{\gamma}{\beta} \rceil, n]$. Signatory i has neither a material nor a reciprocity incentive to deviate to not investing at h. Non-signatory i faces identical incentives to the full investment profile in $\widehat{\Gamma}_P$, thus i does not deviate to not investing at h if $Y \geq Y^*$.

Now consider $h = a^1$ such that $m(a^1) = \lceil \frac{\gamma}{\beta} \rceil$. Signatory i has neither a material nor a reciprocity incentive to deviate to not investing at h. Non-signatory i has a material incentive to deviate to not investing at h and a reciprocity incentive to not do so. Reason as follows to identify Y such that i

does not deviate. Non-signatory i's increase in reciprocity payoff from playing s_i^* rather than $s_i'(h, s_i^*)$ is (4). This is no less than the reduction in i 's material payoff from playing s_i^* rather than $s_i'(h, s_i^*)$ if $Y \geq \overline{Y}(n, \beta, \gamma, m(a^1))$, where

$$\overline{Y}\left(n, \beta, \gamma, m\left(a^1\right)\right) := \frac{2(\gamma - \beta)}{\beta\left(\left(n\beta - \gamma - \beta\right) m\left(a^1\right) + (n-1)\beta\right)}.$$

Note that $\overline{Y}(n, \beta, \gamma, m(a^1)) \geq Y^*$ iff $\frac{\gamma}{\beta} + 1 \geq n$, however by assumption $\left\lceil \frac{\gamma}{\beta} \right\rceil < n$, therefore $\overline{Y}(n, \beta, \gamma, m(a^1)) < Y^*$. Thus $Y \geq Y^*$ is sufficient to prevent non-signatory i deviating at h.

Now consider $h = a^1$ such that $m(a^1) = \left\lceil \frac{\gamma}{\beta} \right\rceil - q$ where $q > 0$ and even, so all players invest. Non-signatory i, faces identical incentives as a non-signatory at a history with $\left\lceil \frac{\gamma}{\beta} \right\rceil$ signatories, thus i does not deviate to not investing if $Y \geq \overline{Y}(n, \beta, \gamma, m(a^1))$, which holds for $Y \geq Y^*$.

Signatory i has a material incentive to deviate to not investing. Using (5), signatory i's change in reciprocity payoff from playing s_i^* rather than $s_i'(h, s_i^*)$ is strictly positive iff $\frac{Y\Omega}{2} > 0$ where

$$\Omega := \left(m\left(a^1\right) - 1\right)\left(\beta - \frac{\gamma}{m\left(a^1\right)}\right)\Delta\lambda_S + \left(n - m\left(a^1\right)\right)\beta\Delta\lambda_N,$$

$\Delta\lambda_S := (n+1)\beta - \frac{m(a^1)+1}{m(a^1)}\gamma$ and $\Delta\lambda_N := (n-1)\beta - \frac{m(a^1)}{m(a^1)+1}\gamma$. Note that $\Delta\lambda_S < \Delta\lambda_N$. To determine the sign of Ω reason as follows. Clearly $m(a^1) - 1 > 0$, $\beta - \frac{\gamma}{m(a^1)} < 0$ and $n - m(a^1) > 0$, thus we need only sign $\Delta\lambda_S$ and $\Delta\lambda_N$. Given $n > \left\lceil \frac{\gamma}{\beta} \right\rceil$, it follows that $(n-1)\beta > \gamma$, which implies that $\Delta\lambda_N > 0$. Consider how $\Delta\lambda_S$ influences the sign of Ω. If $\Delta\lambda_S \leq 0$, then $\Omega > 0$. If $\Delta\lambda_S > 0$, since $\Delta\lambda_S < \Delta\lambda_N$ we can write

$$\Omega > \left(\left(m\left(a^1\right) - 1\right)\left(\beta - \frac{\gamma}{m\left(a^1\right)}\right) + \left(n - m\left(a^1\right)\right)\beta\right)\Delta\lambda_S > 0.$$

where the final inequality follows from $(m(a^1) - 1)(\beta - \frac{\gamma}{m(a^1)}) + (n - m(a^1))\beta > \Delta\lambda_N > 0$. Therefore $\Omega > 0$ and signatory i's reciprocity payoff is strictly higher playing s_i^* instead of $s_i'(h, s_i^*)$. This increase in reciprocity payoff is no less than i's reduction in material payoff from playing s_i^* rather than $s_i'(h, s_i^*)$ if $Y \geq \widehat{Y}(n, \beta, \gamma, m(a^1))$ where

$$\widehat{Y}\left(n, \beta, \gamma, m\left(a^1\right)\right) := \frac{2\left(\gamma / m\left(a^1\right) - \beta\right)}{\Omega}.$$

Now argue that $\widehat{Y}(n, \beta, \gamma, m(a^1)) < Y^*$. To do so, take a function, $\tilde{Y}(n, \beta, \gamma, m(a^1))$, such that $\tilde{Y}(n, \beta, \gamma, m(a^1)) > \widehat{Y}(n, \beta, \gamma, m(a^1))$. To identify an appropriate function, reason as follows. For a given $\Delta\lambda_N$, Ω is decreasing in $\Delta\lambda_S$, and $\Delta\lambda_S$ is bounded by $\Delta\lambda_N$, thus let $\Delta\lambda_S = \Delta\lambda_N$ to minimise Ω. Furthermore, note that Ω is increasing in $\Delta\lambda_N$, and that $\Delta\lambda_N$ is strictly greater than β. To see this, note that $\beta n - \gamma > \beta$ since $n > \frac{\gamma}{\beta} + 1$ by assumption. Also note that $\frac{\gamma}{m(a^1)+1} - \beta > 0$ for all $m(a^1) \in \{1, \ldots, \lceil \frac{\gamma}{\beta} \rceil - 2\}$. Putting this together gives $\Delta\lambda_N > \beta$. Overall then, substitute $\Delta\lambda_S = \Delta\lambda_N = \beta$ into $\widehat{Y}(n, \beta, \gamma, m(a^1))$ to give

$$\tilde{Y}\left(n, \beta, \gamma, m\left(a^1\right)\right) := \frac{2(\frac{\gamma}{m(a^1)} - \beta)}{\beta((n-1)\beta - \frac{m(a^1)-1}{m(a^1)}\gamma)}.$$

Suppose that $\tilde{Y}(n, \beta, \gamma, m(a^1)) > Y^*$. This requires

$$(\gamma - \beta)\left((n-1)\beta - \frac{m(a^1)-1}{m(a^1)}\gamma\right) < \left(\frac{\gamma}{m(a^1)} - \beta\right)\beta(n-1).$$

Note that the LHS is increasing in $m(a^1)$ and that the RHS is decreasing in $m(a^1)$. Substituting $m(a^1) = 1$ gives $(\gamma - \beta)(n-1)\beta < (\gamma - \beta)(n-1)\beta$, a contradiction. Therefore $\tilde{Y}(n, \beta, \gamma, m(a^1)) \leq Y^*$. Overall, $Y^* \geq \tilde{Y}(n, \beta, \gamma, m(a^1)) > \widehat{Y}(n, \beta, \gamma, m(a^1))$, thus $Y \geq Y^*$ is sufficient to prevent signatory i deviating from s_i^* to $s_i'(h, s_i^*)$ at h.

Now consider $h = a^1$ such that $m(a^1) = \lceil \frac{\gamma}{\beta} \rceil - q$ where $q > 1$ and odd, so zero players invest. Non-signatory i has neither material nor reciprocity incentives to deviate to investing at h. Signatory i has no material incentive to deviate to investing at h. Using (5), the change in signatory i's reciprocity payoff from playing s_i^* rather than $s_i'(h, s_i^*)$ is $Y\Omega/2$, which is strictly positive as already established, thus i does not deviate at h.

Now consider $h = a^1$ such that $m(a^1) = \lceil \frac{\gamma}{\beta} \rceil - 1$, so zero invest. Non-signatory i has neither a material nor a reciprocity incentive to deviate to investing. Signatory i has no material incentive to deviate to investing. Using (5), the change in signatory i's reciprocity payoff from playing s_i^* rather than $s_i'(h, s_i^*)$ is $\frac{Y}{2}(\Omega + (n - m(a^1))\beta(\beta - \gamma/(m(a^1)+1)))$, which is strictly positive as we know $\Omega > 0$ and that $\frac{\gamma}{\beta} < m(a^1) + 1$, implies $\frac{\beta - \gamma}{m(a^1)+1} > 0$, thus i does not deviate at h.

Now consider $h = a^1$ such that $m(a^1) = 0$, so zero invest for $\lceil \frac{\gamma}{\beta} \rceil$ odd and n invest otherwise. For $\lceil \frac{\gamma}{\beta} \rceil$ odd, player i has neither material nor reciprocity incentives to deviate to investing. For $\lceil\frac{\gamma}{\beta}\rceil$ even, player i faces identical incentives as a non-signatory following a history of $\lceil\frac{\gamma}{\beta}\rceil$ signatories, thus does not deviate if $Y \geq \overline{Y}(n, \beta, \gamma, m(a^1))$, which is satisfied for all $Y \geq Y^*$.

Finally consider the initial node. Player i has neither material nor reciprocity incentives to deviate. Hence s^* is a SRE. ■

Proof of Proposition 4 (Low reciprocity)

(a) Consider s^* such that $m(Y)$ sign, then i invests iff i signed and there are at least $m(Y)$ signatories. For β sufficiently high, we first identify non-deviation conditions for signatories in the investment stage, then do the same for non-signatories, and then consider the sign-up stage. Using these conditions, we show that for all $Y \in (0, Y^*)$, some s^* is a SRE.

Consider $h = a^1$ such that $m(a^1) > m(Y)$. Signatory i has no material incentive to deviate to not investing. Using (5), the change in i's reciprocity payoff from playing s_i^* rather than $s_i'(h, s_i^*)$ is

$$Y\frac{\beta}{2}\left[(\beta - \frac{\gamma}{m(a^1)})(m(a^1) - 1) - \beta(n - m(a^1))\right]. \tag{6}$$

Note that if $m(a^1) = \frac{\gamma}{\beta}$ then (6) < 0, if $m(a^1) = n$ then (6) > 0 and that (6) is strictly increasing in $m(a^1)$. There must then exist some $\widetilde{m} \in (\frac{\gamma}{\beta}, n)$ such that if $m(a^1) = \widetilde{m}$ then (6) $= 0$. For $m(a^1) \geq \widetilde{m}$, signatory i does not deviate at h. For $m(a^1) \in [m(Y)+1, \widetilde{m})$, signatory i does not deviate to not investing at h if $Y \leq Y_1(m(a^1))$, where

$$Y_1(m(a^1)) \equiv \frac{\beta - \frac{\gamma}{m(a^1)}}{-\frac{\beta}{2}\left[(\beta - \frac{\gamma}{m(a^1)})(m(a^1) - 1) - \beta(n - m(a^1))\right]}.$$

As $Y(m(a^1))$ is strictly increasing in $m(a^1)$, $Y \leq Y_1(m(Y)+1)$ is a sufficient condition for signatory i to not deviate to not investing at h.

Consider $h = a^1$ such that $m(a^1) = m(Y)$. Signatory i has a material incentive to not deviate to not investing. Using (5), the change in i's reciprocity incentive from playing s_i^* rather than $s_i'(h, s_i^*)$ is

$$f(m(a^1)) \equiv \begin{cases} (\beta - \frac{\gamma}{m(a^1)})\frac{m(a^1)\beta - \gamma}{2}(m(a^1) - 1) - \frac{\beta^2}{2}(n - m(a^1)) & \text{if } m(a^1) \geq 3, \\ (\beta - \frac{\gamma}{m(a^1)})(m(a^1)\beta - \frac{\beta}{2} - \gamma)(m(a^1) - 1) - \frac{\beta^2}{2}(n - m(a^1)) & \text{if } m(a^1) = 2. \end{cases}$$

Note that $f(m(a^1))$ is strictly increasing in $m(a^1)$. If $f(m(a^1)) \geq 0$, signatory i does not deviate to not investing at h. If $f(m(a^1)) < 0$, signatory i does not deviate to not investing at h if $Y \leq Y_2(m(a^1))$, where

$$Y_2(m(a^1)) \equiv \frac{\beta - \frac{\gamma}{m(a^1)}}{-f(m(a^1))}.$$

If $m(Y) \geq 3$, then for $m(a^1) \geq m(Y)$, signatory i does not deviate to not investing if $Y \leq min\{Y_1(m(Y)+1), Y_2(m(Y)\}$. This can be rewritten as $Y \leq Y_1(m(Y))$ since for $m(a^1) \geq 3$, $Y_1(m(a^1)) < Y_2(m(a^1))$ and both are strictly increasing in $m(a^1)$. If $m(Y) = 2$, then for $m(a^1) \geq 2$, signatory i does not deviate to not investing if $Y \leq Y_2(m(Y))$ (since $Y_2(m(a^1)) < Y_1(m(a^1)) < Y_1(m(a^1)+1)$).

Consider $h = a^1$ such that $m(a^1) = m(Y) - 1 \geq 2$. Signatory i has a material incentive to deviate to investing. Using (5), the change in i's reciprocity payoff from playing s_i^* rather than $s_i'(h, s_i^*)$ is

$$Y\frac{\beta}{2}\left[(n - m(a^1))((m(a^1)+1)\beta - \gamma) - (m(a^1)-1)(\frac{\gamma}{m(a^1)} - \beta)\right]. \tag{7}$$

Note that (7) > 0. Thus signatory i does not deviate to not investing at h if $Y \geq Y_3(m(a^1))$, where

$$Y_3(m(a^1)) \equiv \frac{\beta - \frac{\gamma}{m(a^1)}}{\frac{\beta}{2}\left[(n - m(a^1))((m(a^1)+1)\beta - \gamma) - (m(a^1)-1)(\frac{\gamma}{m(a^1)} - \beta)\right]}.$$

Consider $h = a^1$ such that $m(a^1) \in [2, m(Y)-1)$. Signatory i has a material incentive to deviate to investing. Using (5), the change in i's reciprocity payoff from playing s_i^* rather than $s_i'(h, s_i^*)$ is

$$Y\frac{\beta}{2}\left[(n - m(a^1))\beta - (m(a^1)-1)(\frac{\gamma}{m(a^1)} - \beta)\right]. \tag{8}$$

Note that (8) > 0. Thus signatory i does not deviate to not investing at h if $Y \geq Y_4(m(a^1))$, where

$$Y_4(m(a^1)) \equiv \frac{\beta - \gamma/m(a^1)}{\frac{\beta}{2}\left[(n - m(a^1))\beta - (m(a^1)-1)(\frac{\gamma}{m(a^1)} - \beta)\right]}.$$

Since $Y_4(.)$ is increasing in $m(a^1)$, for all $m(a^1) \in [2, m(Y)-1)$, signatory i does not deviate if $Y \geq Y_4(m(Y)-2)$.

Consider $h = a^1$ such that $m(a^1) = 1$. Signatory i has no material incentive to deviate to investing. Using (5), the change in i's reciprocity payoff from playing s_i^* rather than $s_i'(h, s_i^*)$ is $Y\frac{\beta^2}{2}(n-1) > 0$, thus i has no reciprocity incentive to deviate either. Finally, Consider $h = a^1$ such that $m(a^1\ ) \in \{0, 1\}$, i has neither material nor reciprocity incentive to deviate to investing.

Now consider non-signatories. Clearly non-signatory i has no material incentive to invest. For all $h = a^1$ such that $m(a^1) \notin \{m(Y)-1, m(Y)\}$, non-signatory i perceives others as no more kind than in the full investment profile in $\widehat{\Gamma}_P$, thus does not deviate to investing for $Y < Y^*$. For $h = a^1$ such that $m(a^1) = m(Y)-1$, all other players are unkind to i so he has no reciprocity incentive to deviate to investing. For $h = a^1$ such that $m(a^1) = m(Y)$, using (4), non-signatory i's reciprocity incentive from deviating to investing is $\frac{\beta^2}{2}(n-(m(a^1))^2 - m(a^1) - 1)$, which is non-negative if $n \geq 7$.

Now consider the sign-up stage. Signatory i has identical incentives to a signatory's incentives at $h = a^1$ such that $m(a^1) = m(Y)$. Non-signatory i has identical incentives to a non-signatory's incentives at $h = a^1$ such that $m(a^1) = m(Y)$. Thus if players have no incentive to deviate in the investment stage, they also have no incentive to deviate in the sign-up stage.

In sum, there exists a SRE where $m(Y) = 2$ if $Y \in I_2 \equiv \{Y : Y < Y_2(m(Y))\}$; $m(Y) = 3$ if $Y \in I_3 \equiv \{Y : Y \in [Y_3(m(Y)-1), Y_1(m(Y))]\}$; $m(Y) \in [4, \widetilde{m})$ if $Y \in I_{[4,\widetilde{m})} \equiv \{Y : Y \in [Y_4(m(Y)-1), Y_1(m(Y))]\}$ and $m(Y) \geq \widetilde{m}$ if $Y \in I_{\geq \widetilde{m}} \equiv \{Y : Y \geq Y_4(m(Y)-1)\}$. To show that there exists a SRE for all $Y \in (0, Y^*)$, verify that $I_2 \cup I_3 \cup I_{[4,\widetilde{m})} \cup I_{\geq \widetilde{m}}$ covers $\mathbb{R}^+$ as follows. First, $I_2 \cap I_3 \neq \varnothing$ as $Y_3(m(Y)-1) < Y_2(m(Y)-1) < Y_2(m(Y))$. Second, $I_3 \neq \varnothing$ since $Y_3(m(Y)-1) < Y_1(m(Y)-1) < Y_1(m(Y))$. Third, since $Y_4(m(Y)) < Y_1(m(Y))$ and both increase in $m(Y)$, it holds that $Y_4(m(Y)-1) < Y_4(m(Y)) < Y_1(m(Y)) < Y_1(m(Y)+1)$. Therefore the intersections for the intervals for all $m(Y) \geq 3$ are also non-empty.

(b) See final paragraph of part (a) of this proof. ■

References

[1] Andreoni, J. and H. Varian (1999) "Preplay contracting in the prisoners' dilemma" *Proceedings of the National Academy of Science* 96: 10933-10938.

[2] Barrett, S. (1994) "Self-enforcing international environmental agreements" *Oxford Economic Papers* 46: 878-894.

[3] Barrett, S. (2003) *Environment and Statecraft: The Strategy of Environmental Treaty-Making*. Oxford. Oxford University Press.

[4] Battaglini, M. and B. Harstad (2016) "Participation and the duration of environmental agreements" *Journal of Political Economy* 124(1): 160-204.

[5] Battigalli, P. and M. Dufwenberg (2009) "Dynamic psychological games" *Journal of Economic Theory* 144: 1-35.

[6] Bierbrauer, F. and N. Netzer (2016) "Mechanism design and intentions" *Journal of Economic Theory* 163: 557-603.

[7] Bierbrauer, F., Ockenfels, A., Pollak, A. and D. Rückert (2015) "Robust mechanism design and social preferences" mimeo.

[8] Boadway, R., Song, Z. and J.-F. Tremblay (2007) "Commitment and matching contributions to public goods" *Journal of Public Economics* 91: 1664-1683.

[9] Charness, G., Fréchette, G. R. and C.-Z. Qin (2007) "Endogenous transfers in the prisoner's dilemma game: An experimental test of cooperation and coordination" *Games and Economic Behavior* 60: 287-306.

[10] Charness, G. and M. Rabin (2002) "Understanding social preferences with simple tests" *Quarterly Journal of Economics* 117: 817-869.

[11] Clémençon, R. (2016) "The two sides of the Paris climate agreement" *Journal of Environment & Development* 25(1): 3-24.

[12] Dufwenberg, M. and G. Kirchsteiger (2004) "A theory of sequential reciprocity" *Games and Economic Behavior* 47: 268-298.

[13] Dufwenberg, M. and A. Patel (2016) "Reciprocity networks and the participation problem" *Games and Economic Behavior*, forthcoming.

[14] Ellingsen, T. and E. Paltseva (2016) "Confining the Coase theorem: contracting, ownership and free-riding" *Review of Economic Studies* 83: 547-586.

[15] Falk, A. and U. Fischbacher (2006) "A theory of reciprocity" *Games and Economic Behavior* 54: 293-315.

[16] Falkinger, J., Fehr, E., Gächter, S. and R. Winter-Ebmer (2000) "A simple mechanism for the efficient provision of public goods: Experimental evidence" *American Economic Review* 90: 247-264.

[17] Fischbacher, U., Fehr, E. and S. Gächter (2001) "Are people conditionally cooperative? Evidence from a public goods experiment" *Economic Letters* 71: 397-404.

[18] Geanakoplos, J., Pearce, D. and E. Stacchetti (1989) "Psychological games and sequential rationality" *Games and Economic Behavior* 1: 60-79.

[19] Greenstone, M. (2015) "Suprisingly, a voluntary climate agreement could actually work" *New York Times*, The Upshot, February 13th.

[20] Guttman, J. (1978) "Understanding collective action: Matching behavior" *American Economic Review* 68: 251-255.

[21] Guttman, J. (1987) "A non-Cournot model of voluntary collective action" *Economica* 54: 1-19.

[22] Hadjiyiannis, C., Iris, D. and C. Tabakis (2012) "International environmental cooperation under fairness and reciprocity" *B.E. Journal of Economic Analysis & Policy (Topics)* 12(1): Article 33.

[23] Harstad, B. (2012) "Climate contracts: A game of emissions, investments, negotiations and renegotiations" *Review of Economic Studies* 79: 1527-1557.

[24] Harstad, B. (2015) "The dynamics of climate agreements" *Journal of the European Economic Association* 14(3): 719-752.

[25] Jackson, M. O. and S. Wilkie (2005) "Endogenous games and mechanisms: Side payments among players" *Review of Economic Studies* 72: 543-566.

[26] Jang, D. (2015) "Reciprocity and International Environmental Agreements" in *Two Essays of Other-Regarding Preferences on Social Decision Making*, Chapter 2, PhD thesis. University of Arizona.

[27] Jaquet, J. and D. Jamieson (2016) "Soft but significant powers in Paris agreement" *Nature Climate Change* 6: 643-646.

[28] Katz, M. L. (1986) "An analysis of cooperative R&D" *RAND Journal of Economics* 17(4): 527-543.

[29] Kolstad, C. D. (2014) "International Environmental Agreements among heterogenous countries with social preferences" *NBER Working Paper* No. 20204.

[30] Martimort, D. and W. Sand-Zantman (2016) "A mechanism design approach to climate-change agreements" *Journal of the European Economic Association* 14(3): 669-718.

[31] Netzer, N. and A. Volk (2014) "Intentions and ex-post implementation" mimeo.

[32] Nyborg, K. (2015) "Reciprocal climate negotiators" IZA DP no. 8866.

[33] Rabin, M. (1993) "Incorporating fairness into game theory and economics" *American Economic Review* 83: 1281-1302.

[34] Sugden, R. (1984) "Reciprocity: the supply of public goods through voluntary contributions" *Economic Journal* 94: 772-787.

[35] Varian, H. (1994) "A solution to the problem of externalities when agents are well-informed" *American Economic Review* 84: 1278-1293.