



Reciprocity networks and the participation problem [☆]



Martin Dufwenberg ^{a,b,c,d}, Amrish Patel ^{b,*}

^a Bocconi University, IGIER, Italy

^b University of Gothenburg, Sweden

^c University of Arizona, United States

^d CESifo, Germany

ARTICLE INFO

Article history:

Received 3 July 2014

Available online 3 September 2015

JEL classification:

C72

D03

H41

Keywords:

Discrete public good

Participation

Reciprocity networks

Coordination

Cost-sharing

ABSTRACT

Reciprocity can be a powerful motivation for human behaviour. Scholars have argued that it is relevant in the context of private provision of public goods. We examine whether reciprocity can resolve the associated coordination problem. The interaction of reciprocity with cost-sharing is critical. Neither cost-sharing nor reciprocity in isolation can solve the problem, but together they have that potential. We introduce new network notions of reciprocity relations to better understand this.

© 2015 Elsevier Inc. All rights reserved.

1. Introduction

You receive an email from your boss stating that he must nominate exactly two staff as trade union representatives as soon as possible. All staff would like that someone represents them on the union, but the question is who? If no-one else volunteers, or if at least two others volunteer, there is no sense in you volunteering. If only one person volunteers then the outcome depends on you. You click reply and hesitate... how do your colleagues choose?

The story illustrates the participation problem in the private provision of discrete public goods. Grander scale examples might include international economic agreements, open source software, and political representation. A useful framework for analysing the issue is found in the seminal work of Palfrey and Rosenthal (1984) (P&R). They identify two types of equilibria, that where no-one participates and that where exactly those needed for provision do so. This dichotomy exposes two fundamental problems, concerning efficiency (since all individuals would be better off with provision than without) and coordination (since there are multiple provision equilibria). The problems can neither be solved by evoking the refinement of strict equilibrium nor by using refunds, that is returning costs if there are insufficient participants. However, P&R show that taken together, these two factors can achieve efficiency: with refunds, all strict equilibria are efficient.

[☆] We thank Yann Bramoullé, Vince Crawford, David Gill, Alex Sebald, Bob Sugden, an associate editor, two anonymous referees and several seminar audiences for many useful comments. Phil Woods provided able research assistance. Patel acknowledges generous funding from the Jan Wallander and Tom Hedelius Foundation Browaldh Research Fellowship (I2012-0091-1) and project grant (P2012-0097-1) and the hospitality of the Department of Economics, University of Oxford. Dufwenberg acknowledges the financial support of ERC grant 324219 ("Stratemotions").

* Corresponding author.

E-mail addresses: martin.dufwenberg@unibocconi.it (M. Dufwenberg), amrish.patel@economics.gu.se (A. Patel).

These important insights presume that individuals maximise material payoffs. But scholars argue that in the context of public goods provision it is natural for additional motives such as reciprocity to come into play (Sugden, 1984).¹ Reciprocity can be modelled as the desire to be kind to those who are kind to you and unkind to those who are unkind to you (Rabin, 1993; Dufwenberg and Kirchsteiger, 2004 (D&K), Falk and Fischbacher, 2006). In settings where it is beneficial that everyone contributes, the application of reciprocity theory is straightforward and leads to conditional cooperation: players desire to contribute increases as others' contributions go up. Empirical support for such behaviour has been provided (Keser and van Winden, 2000; Fischbacher et al., 2001; Fischbacher and Gächter, 2010) via experimental public good games (Ledyard, 1995). However, in discrete public goods games it isn't useful, but rather wasteful, to contribute if many others do so anyway, so the implications of reciprocity are not obvious. Will reciprocity make it easier to decide how to reply to your boss' email? Will it solve the efficiency and coordination problems? Answers require systematic analysis.

We use the workhorse model of P&R and conceptualise reciprocity as in the D&K model. Cost-sharing, where individuals reduce each others' cost-burden if the number of participants exceeds that needed for provision, proves crucial. P&R's insights are robust to reciprocity and cost-sharing, taken individually. However, with both reciprocity and cost-sharing in the picture, conclusions are dramatically different. There may exist a unique efficient refined equilibrium so the coordination problem can be solved. But results on this potential are marvellously intricate, with possibilities as well as dead ends in the form of multiplicity or non-existence.

To untangle structure, we introduce and study new network notions of reciprocity relationships that describe the players' attitudes towards each other. Methodologically, we are thus connected to an infant literature on games with networks of social preferences (Leider et al., 2009; Bourlès et al., 2015). Among other things, our approach demonstrates the coordinating power of "reciprocity cliques" in a world of non-reciprocal players and how "reciprocity alienation" can impede coordination.

Our uniqueness results may seem surprising for those who believe that "in standard examples, the notion of reciprocal preferences tends to increase the number of equilibria" (Sobel, 2005, p. 410). We examine an economically important class of games that under standard preferences seems intrinsically plagued with multiple equilibria, yet reciprocity potentially offers a solution.

Section 2 recaps P&R and adds cost-sharing (2.1), then defines and highlights important properties of reciprocity preferences (2.2 and 2.3). Section 3 presents results, on reciprocity in P&R (3.1) and on reciprocity and cost-sharing (3.2–3.3). Section 4 examines welfare (4.1) and other social preferences (4.2), before we conclude in Section 5.

2. Model

2.1. P&R and cost-sharing

We first recall P&R's classic participation game. Let $N = \{1, 2, \dots, n\}$ be the set of players, with $n \geq 3$. Player i 's strategy set is $S_i = \{0, 1\}$ where 1 corresponds to participating in provision and 0 to not doing so. We focus on pure strategies throughout. Let $S = \times_{i \in N} S_i$. Write $s = (s_1, s_2, \dots, s_n)$ for a profile of strategies.

The threshold number of participants for provision of the public good is w , where $1 < w < n$. Let the cost of participation be $x \in (0, 1)$ and each player receives an additional payoff of 1 if the good is provided. Let π_i be player i 's material payoff function.

P&R introduce two variants of their game differing in whether or not costs are refunded when there are fewer than w participants. To define payoff functions, let m denote the number of players that participate. With no refunds, player i 's material payoff function is defined by

$$\text{No Refunds: } \pi_i(s) = \begin{cases} 1 & \text{if } m \geq w \text{ and } s_i = 0, \\ 1 - x & \text{if } m \geq w \text{ and } s_i = 1, \\ 0 & \text{if } m < w \text{ and } s_i = 0, \\ -x & \text{if } m < w \text{ and } s_i = 1. \end{cases} \tag{1}$$

Any profile where $m = 0$ or $m = w$ is a Nash equilibrium, and in fact also a strict equilibrium.

P&R show that refunds imply the inefficient no participation outcome is no longer a strict equilibrium. With refunds, i 's material payoff function is defined by

$$\text{Refunds: } \pi_i(s) = \begin{cases} 1 & \text{if } m \geq w \text{ and } s_i = 0, \\ 1 - x & \text{if } m \geq w \text{ and } s_i = 1, \\ 0 & \text{if } m < w. \end{cases} \tag{2}$$

While any profile where $m = w$ remains a strict equilibrium, $m = 0$ does not because deviation to participation would be costless. All strict equilibria are thus efficient.

¹ Reciprocity is not the only important social preference in such environments. Makris (2009) considers altruism and Rothenhäusler et al. (2013) a form of guilt.

We will additionally consider a counterpart to refunds for $m > w$, cost-sharing, where costs in excess of the provision cost are returned in equal shares to the participating players when $m > w$. Player i 's material payoff function is then defined by

$$\text{Cost-sharing: } \pi_i(s) = \begin{cases} 1 & \text{if } m \geq w \text{ and } s_i = 0, \\ 1 - \frac{w}{m}x & \text{if } m \geq w \text{ and } s_i = 1, \\ 0 & \text{if } m < w. \end{cases} \tag{3}$$

In some contexts cost-sharing is very natural. For instance, in our opening example 2 workers were needed to be union representatives. Suppose that 4 workers volunteer. Distributing the union tasks between the 4 seems more reasonable than making all 4 work on every task.

If players are motivated by material payoffs alone, pure strategy Nash and strict equilibria are unaffected by cost-sharing as it only affects payoffs when $m > w$ and such profiles are neither equilibria nor attractive profiles to deviate to given the waste involved.² However with reciprocity motivation cost-sharing will matter.

2.2. Reciprocity: definitions

We next incorporate preferences for reciprocity. Following D&K,³ when player i plays strategy s_i and holds (point) belief $(b_{ij})_{j \neq i} \in \prod_{j \neq i} S_j$ about other players' strategies, i 's kindness to player j is

$$\kappa_{ij}(s_i, (b_{ij})_{j \neq i}) = \pi_j(s_i, (b_{ij})_{j \neq i}) - \frac{1}{2} \left[\max_{s'_i} \pi_j(s'_i, (b_{ij})_{j \neq i}) + \min_{s'_i} \pi_j(s'_i, (b_{ij})_{j \neq i}) \right]. \tag{4}$$

The first term on the right-hand side of (4) is the material payoff that i believes that j receives. To define kindness, this payoff must be compared to some reference point. The reference point is the second term, the average of the highest and the lowest material payoff i believes he could "give" to j . If $\kappa_{ij}(\cdot) > 0$, i is kind to j , if $\kappa_{ij}(\cdot) < 0$, i is unkind to j and if $\kappa_{ij}(\cdot) = 0$, i has zero kindness to j .

To illustrate, take the game with no refunds (material payoff function (1)). Let $n = 3$, $w = 2$ and $x = \frac{1}{2}$. Suppose that player 1 believes that only player 3 is participating ($b_{12} = 0$, $b_{13} = 1$). Player 1 is then kind to both players if he participates ($\kappa_{12}(1, (0, 1)) = 1 - \frac{1}{2}(0 + 1) = \frac{1}{2}$; $\kappa_{13}(1, (0, 1)) = \frac{1}{2} - \frac{1}{2}(-\frac{1}{2} + \frac{1}{2}) = \frac{1}{2}$) and unkind to both players if he does not ($\kappa_{12}(0, (0, 1)) = 0 - \frac{1}{2}(0 + 1) = -\frac{1}{2}$; $\kappa_{13}(0, (0, 1)) = -\frac{1}{2} - \frac{1}{2}(-\frac{1}{2} + \frac{1}{2}) = -\frac{1}{2}$).

Analogously to kindness, when i holds (point) belief b_{ij} about j 's strategy and (point) belief $(c_{ijk})_{k \neq j} \in \prod_{k \neq j} S_k$ about j 's belief other players' strategies, i 's perceived kindness of j towards i is

$$\lambda_{iji}(b_{ij}, (c_{ijk})_{k \neq j}) = \pi_i(b_{ij}, (c_{ijk})_{k \neq j}) - \frac{1}{2} \left[\max_{b'_{ij}} \pi_i(b'_{ij}, (c_{ijk})_{k \neq j}) + \min_{b'_{ij}} \pi_i(b'_{ij}, (c_{ijk})_{k \neq j}) \right]. \tag{5}$$

The first term on the right-hand side of (5) is the material payoff that i believes that j believes that i receives. The reference point is the average of the highest and the lowest material payoff that i believes that j believes that he could give to i . If $\lambda_{iji}(\cdot) > 0$, i perceives that j is kind to him... etc.

Player i 's utility is the sum of his material and reciprocity payoffs:

$$U_i(s, (b_{ij}, (c_{ijk})_{k \neq j})_{j \neq i}) = \underbrace{\pi_i(s)}_{\text{material payoff}} + \underbrace{\sum_{j \in N \setminus \{i\}} (Y_{ij} \cdot \kappa_{ij}(s_i, (b_{ij})_{j \neq i}) \cdot \lambda_{iji}(b_{ij}, (c_{ijk})_{k \neq j}))}_{\text{reciprocity payoff}}, \tag{6}$$

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

The assumption that Y_{ij} is common knowledge is clearly restrictive. It does however make sense with a small group of familiar players, for example a work team or the set of chief politicians. It may also be approximately true for people who

² It would however affect Nash equilibria in mixed strategies, as allowed for in P&R, and also has implications in related games. For example, Shinohara (2009) where players have heterogeneous valuations of the public good and Makris (2009) where there is incomplete information over whether or not players are altruistic.

³ D&K's theory of reciprocity applies to a rather general class of extensive game forms. Our focus on a specific class of games affords us two simplifications: (i) we need not condition strategies and beliefs on histories as the game has simultaneous moves; (ii) we need not discuss D&K's notion of "inefficient strategies" (p. 276) as there are no such strategies.

self-select into a group. Might people who work for a charity have similar Y_{ij} ? Common knowledge is perhaps more likely when preferences are very similar. Extending the model to asymmetric information is beyond our scope, but an important avenue for future work.

Finally, we state the two solution concepts that we use. Note that we may write $s = (s_i, s_{-i})$ where s_{-i} is a profile of strategies for all players except i .

Definition 1 (SRE). Profile s is a sequential reciprocity equilibrium (SRE) if for all $i \in N$

1. $U_i \left(s, \left(b_{ij}, (c_{ijk})_{k \neq j} \right)_{j \neq i} \right) \geq U_i \left(s'_i, s_{-i}, \left(b_{ij}, (c_{ijk})_{k \neq j} \right)_{j \neq i} \right)$ for all s'_i ,
2. $c_{ijk} = b_{jk} = s_k$ for all $j \neq i$ and $k \neq j$.

The first condition requires that i is best-responding given others' strategies and his beliefs. The second demands that players hold correct beliefs. If for all $i, j \in N, Y_{ij} = 0$, then Definition 1 describes a Nash equilibrium (+ correct beliefs) in a game where utility equals material payoffs.

While a unilateral deviation from a SRE cannot increase the deviator's payoff, such a deviation from a SSRE leads to an actual loss:

Definition 2 (SSRE). Profile s is a strict sequential reciprocity equilibrium (SSRE) if for all $i \in N$

1. $U_i \left(s, \left(b_{ij}, (c_{ijk})_{k \neq j} \right)_{j \neq i} \right) > U_i \left(s'_i, s_{-i}, \left(b_{ij}, (c_{ijk})_{k \neq j} \right)_{j \neq i} \right)$ for all $s'_i \neq s_i$,
2. $c_{ijk} = b_{jk} = s_k$ for all $j \neq i$ and $k \neq j$.

If for all $i, j \in N, Y_{ij} = 0$, Definition 2 describes a strict equilibrium (+ correct beliefs) in a game where utility equals material payoffs.

2.3. Reciprocity: highlighted properties

We now draw attention to properties of reciprocity theory in our model that will help in understanding results.

Psychological game theory

Reciprocity theory (Rabin, 1993, D&K 2004) is developed within the framework of psychological game theory (Geanakoplos et al., 1989; Battigalli and Dufwenberg, 2009), where in contrast to classical game theory utilities depend not only on actions chosen, but potentially also on belief hierarchies. Consider the following example for a motivational dependence that psychological games can capture but classical games cannot.

Example 0. Let $n = 3$ and $w = 2$. Suppose player 1 believes that only player 3 is participating ($b_{12} = 0, b_{13} = 1$). Does player 2 deserve negative reciprocity from player 1? According to D&K this depends on 1's beliefs about 2's beliefs about others' strategies. Contrast when 1 believes 2 believes that only 3 is participating ($c_{121} = 0, c_{123} = 1$), with when 1 believes 2 believes that no-one is participating ($c_{121} = 0, c_{123} = 0$). Only in the former case does 1 believe that 2 believes himself to be pivotal for the good's provision and thus intentionally lowers 1's payoff. By (5), 1 perceives 2 as unkind in the former case ($\lambda_{121}(0, (0, 1)) = 0 - \frac{1}{2}(0 + 1) = -\frac{1}{2}$) and as having zero kindness in the latter ($\lambda_{121}(0, (0, 0)) = 0 - \frac{1}{2}(0 + 0) = 0$). In a psychological game 1 may react differently in these two cases. Indeed, D&K permit negative reciprocity in the former but not in the latter (applying (6)). Classical game theory would not allow distinct reactions since 1 believes the same action profile is played in both cases.

Pivotality and kindness

Reciprocity incentives in Example 0 depend on the perceived ability of player 2 to influence player 1's material payoff, his pivotality.⁴ Non-pivotality will drive many of our results. Payoff functions (1)–(3) imply i is non-pivotal for j 's material payoff if the good will either: (a) not be provided regardless of i 's strategy choice, or; (b) be provided regardless of i 's strategy choice and j is not participating so incurs no costs, or; (c) be provided regardless of i 's strategy choice and j participates but there is no cost-sharing so i cannot reduce j 's cost-share by participating.

Applying cases (a)–(c) to kindness functions (4) and (5), we can establish two lemmas showing that non-pivotality implies zero kindness. These will be useful when stating and proving later results.

If i believes he is non-pivotal for j 's material payoff then both his strategies have a kindness towards j equal to zero, $\kappa_{ij}(\cdot) = 0$ (apply (4)).

⁴ See Bartling et al. (2015) for an experiment where a similar notion of pivotality plays an important role in explaining behaviour.

Lemma 1 (Zero kindness). For some $i, j \in N$, $(b_{ij})_{j \neq i}$ and all $s_i \in S_i$, $\kappa_{ij}(\cdot) = 0$ if and only if i believes that either:

- (a) there are fewer than $w - 1$ participants excluding i , or that,
- (b) there are at least w participants excluding i and j is not participating, or that,
- (c) there are at least w participants excluding i and j is participating, but there is no cost-sharing.

Similarly if i believes that j believes that j is non-pivotal for i 's material payoff, both of j 's strategies have a perceived kindness towards i equal to zero, $\lambda_{iji}(\cdot) = 0$.

Lemma 2 (Zero perceived kindness). For some $i, j \in N$, $(c_{ijk})_{k \neq j}$ and all $b_{ij} \in S_j$, $\lambda_{iji}(\cdot) = 0$ if and only if i believes that j believes that either:

- (a) there are fewer than $w - 1$ participants excluding j , or that,
- (b) there are at least w participants excluding j and i is not participating, or that,
- (c) there are at least w participants excluding j and i is participating but there is no cost-sharing.

It follows from these lemmas that either both strategies have zero (perceived) kindness, or one is (perceived as) kind and the other unkind (see (4) and (5)). It also follows from the lemmas that when beliefs are in equilibrium, $\kappa_{ij}(\cdot) = 0$ if and only if $\lambda_{jij}(\cdot) = 0$.

If for some players i and j either of the lemmas hold, then j offers i no reciprocity incentive (because $\kappa_{ij}(\cdot) \cdot \lambda_{iji}(\cdot) = 0$ in (6)). Furthermore, if this is true for all j , then i 's strategy choice is determined by material payoffs alone even if $Y_{ij} > 0$ for some j (see (6)). It does not however follow that i 's behaviour in SRE (SSRE) is identical to that in NE (strict equilibria) since others' behaviour may still be motivated by reciprocity incentives and thus i 's best response may differ.

3. Results

Does reciprocity matter in discrete public good games? The answer is essentially no in P&R's setting (3.1). The power of reciprocity is unlocked if one considers cost-sharing, as we show first relying on examples (3.2) and then via formal statements that generalise those examples (3.3). All proofs are found in Appendix A; we provide only the main intuition in the text.

3.1. Reciprocity in P&R

The following proposition characterises reciprocity equilibria in P&R's game.

Proposition 1 (Reciprocity in P&R). (i) With no refunds, the set of SRE equals the set of SSRE and is independent of Y_{ij} . (ii) With refunds, the set of SSRE is independent of Y_{ij} .

It follows from the proposition, and the insight that P&R's model may be seen as the special case where for all $i, j \in N$, $Y_{ij} = 0$, that P&R's results are robust to the incorporation of reciprocity. Without refunds there are both equilibrium efficiency and coordination problems; refunds and strictness deal with the efficiency problem. The intuition is as follows:

Without refunds, inefficient no-participation remains an equilibrium despite reciprocity as each player believes he is not pivotal, thus faces no reciprocity incentives (Lemma 1(a)). Furthermore, since deviation incurs a material cost, no participation is also a SSRE.

With refunds, the set of SRE does depend on Y_{ij} .⁵ However, this is of little relevance to us as reasoning analogous to the previous paragraph demonstrates the inefficient SRE still exists. We thus have the same motivation to focus on SSRE in Proposition 1(ii) as P&R had to focus on strict equilibria. To see that inefficient profiles cannot be SSRE, reason as follows. If there are more than w participants, each participant believes he is not pivotal (Lemmas 1(b) and 1(c)) and thus has no reciprocity incentives. Since deviating increases his material payoff, this is not an SSRE. For a profile with a number of participants strictly between 0 and w , each participant again believes he is not pivotal (Lemma 1(a)). This time deviating gives the same material payoff and thus the player incurs no strict loss in utility. Finally, for a no-participation profile, deviation by any does not strictly reduce his utility. All that remains are the efficient profiles, those with w participants.

The central reason why reciprocity has no bite in tackling multiplicity is zero (perceived) kindness due to non-participant non-pivotality in strategy profiles with w participants. To illustrate, let $n = 3$, $w = 2$ and consider a candidate equilibrium where only players 2 and 3 participate. Player 2 is pivotal and equilibrium beliefs imply 1 perceives 2 as kind. However, even if $Y_{12} > 0$, 1 believes he is not pivotal for 2's payoff (Lemma 1(c)) thus cannot reciprocate 2's kindness. Furthermore,

⁵ For those interested, the set of SRE is a superset of that where for all $i, j \in N$, $Y_{ij} = 0$. The only difference arises for profiles where $m = w - 1$ which are SRE if and only if non-participants are sufficiently reciprocal towards one another.

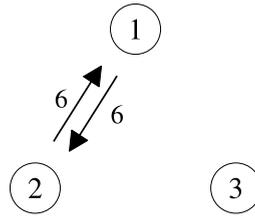


Fig. 1. A reciprocity network.

equilibrium beliefs also mean that 2 acknowledges 1’s perceived non-pivotality (Lemma 2(c)) and so does not treat his non-participation as unkind. Thus 1 provides 2 with no reciprocity incentive to deviate, even if $Y_{21} > 0$. By contrast (anticipating Subsections 3.2–3.3) if there were cost-sharing these latent reciprocity incentives are reignited in such instances by allowing 1 to influence 2’s payoff. This may motivate deviations and gain traction on multiplicity.

3.2. Reciprocity and cost-sharing: examples

Reciprocity has a potentially dramatic effect under cost-sharing. Not only does the set of SSRE now depend on Y_{ij} , but uniqueness becomes a possibility. The coordination problem can thus potentially be solved.

Our results will depend on preferences, specifically the distribution of players’ reciprocity sensitivities, the “reciprocity network”. It is useful to represent this using weighted directed graphs. Vertices represent players and their labels correspond to player labels. A directed edge originating from player i ’s vertex and ending at player j ’s vertex implies $Y_{ij} > 0$. The edge label is Y_{ij} . Fig. 1, for example, represents a 3-player network where for all $i, j \in N$, $Y_{ij} = 0$ except that $Y_{12} = Y_{21} = 6$.

We now define our first network notion of reciprocity and illustrate uniqueness.

Definition 3 (Reciprocity clique of strength ρ). A set of players $L \subseteq N$ is a reciprocity clique of strength ρ if for all $i, j \in L$, $Y_{ij}, Y_{ji} \geq \rho > 0$.

A reciprocity clique may represent a form of friendship.⁶ It seems unlikely in hierarchical relationships such as those between a firm and workers (cf. D&K 2000) but plausible among peers. A sufficiently strong reciprocity clique may overcome the coordination problem as we illustrate in Example 1.

Example 1. Let $n = 3$, $w = 2$ and the reciprocity network be as in Fig. 1. Full participation cannot be a SSRE as player 3 would deviate. Player 3 and just one of the clique members participating cannot be either, since the non-participant clique member’s reciprocity gain from reducing the cost-share of the participating member is greater than his material cost of participation. This leaves only the clique participating as the unique SSRE. The coordination problem is solved.

Compare this to the case without cost-sharing. Only player 3 and one clique member participating would then remain SSRE (see discussion at the end of Section 3.1), so multiplicity would follow.

The existence of a clique is not generally sufficient for uniqueness as Example 2 will show. It also illustrates our second network notion of reciprocity.

Definition 4 (Reciprocity alienation). A set of players $L \subset N$ is reciprocity alienated if for all $i \in L$ and all $j \in N \setminus L$, $Y_{ij} = Y_{ji} = 0$.

Players being reciprocity alienated neither implies nor is implied by them having selfish preferences. For example, players in L having positive reciprocity sensitivities towards one another does not contradict alienation; and, a player not in L with a positive reciprocity sensitivity towards a player in L implies L is not alienated. Intuitively, the notion implies that players inside the alienated group provide no reciprocity incentives to those outside the group, and vice-versa.

Our next example adds a non-reciprocal player to Example 1 to illustrate the coordination problems caused by alienation and that a clique is not sufficient for coordination.

Example 2. Let $n = 4$, $w = 2$ and the reciprocity network be as in Fig. 2. Uniqueness is now impossible despite the reciprocity clique. A SSRE with only 1 and 2 participating exists; however, there also exists one where only 3 and 4 participate. Alienation implies there are no reciprocity incentives acting across the pairs to break the equilibria.

Although both examples have cliques and multiple sets of alienated players, there is uniqueness in the first example and multiplicity in the second. The difference arises because Example 2 has multiple sets of alienated players of size w ,

⁶ Reciprocating unkindness is probably not a particularly common feature of enduring friendships however!

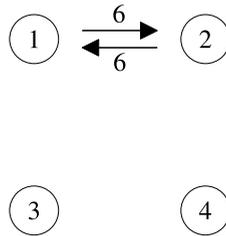


Fig. 2. Reciprocity alienation.

whereas Example 1 does not. Section 3.3 shows that multiple alienated sets of this size prevents uniqueness more generally. Example 3 illustrates how a clique can recover uniqueness.

Example 3. Let $n = 4$, $w = 2$, $x = 0.5$ and the reciprocity network be as in Fig. 3. Players 3 and 4 only participating is now not a SSRE as player 1(2) has an incentive to deviate: the gain from reciprocating the kindness of player 3(4) is greater than the material cost. All profiles except where only the clique participate can be excluded, uniqueness thus re-emerges.

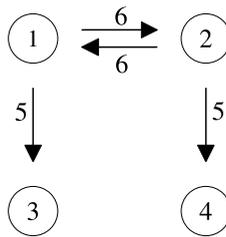


Fig. 3. Preventing alienation.

Example 3 illustrates how a clique being reciprocal towards non-clique players avoids alienation and can imply uniqueness. These precise numbers are not necessary for coordination (see Section 3.3 for uniqueness with general reciprocity networks).

The final example concerns networks with high reciprocity sensitivities. Given that multiple equilibria exist with standard preferences one may conjecture that sufficiently high reciprocity sensitivity solves the coordination problem. This is not necessarily true, however.

Example 4. Let $n = 4$, $x = 0.5$ and the reciprocity network be as in Fig. 4. Consider two cases, first $w = 3$. Three participants cannot be a SSRE as the non-participant deviates since his reciprocity gain from doing so outweighs his material cost. Full participation is the unique SSRE. Now suppose $w = 2$ and 3 players are participating. Participants have no incentive to deviate as their reciprocity gain from reciprocating the kindness of fellow participants is greater than the material cost. Non-participants do not deviate either since equilibrium beliefs imply they perceive a particular participant’s kindness towards them as zero (Lemma 2(b)). Thus, there are multiple SSRE.

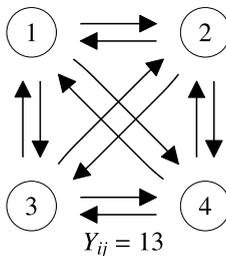


Fig. 4. High reciprocity sensitivity.

More generally high reciprocity gives uniqueness if and only if $w = n - 1$; see Section 3.3. Example 4 also illustrates that there may be more than w participants in a unique SSRE. We show that any number between w and n can be consistent with uniqueness in the next section.

To summarise the insights from [Examples 1–4](#).

1. Existence of a unique SSRE is possible with cost-sharing and reciprocity ([Examples 1, 3 and 4](#)).
2. Reciprocity alienation can prevent uniqueness ([Example 2](#)).
3. A sufficiently strong reciprocity clique with positive reciprocity sensitivities towards players outside the clique can imply a unique SSRE ([Example 3](#)).
4. High reciprocity sensitivities do not necessarily imply uniqueness or multiplicity of SSRE ([Example 4](#)).

Readers looking mainly for key intuitions exhibited by typical examples may now skip to [Section 4](#). Readers interested in seeing some more general formal statements, plus further details, should continue to [Section 3.3](#).

3.3. Reciprocity and cost-sharing: details

First we describe the SSRE of our model. For strategy profile s , let M denote the associated set of participating players and m be the cardinality of M .

Proposition 2 (*Cost-sharing equilibria*). *With cost-sharing, all SSRE involve provision of the public good. Profile s where $m = w$ is a SSRE if and only if for all $i \in M$,*

$$1 - x + \frac{(1 - x)^2}{2} \sum_{j \in M \setminus \{i\}} Y_{ij} > \frac{x}{2(w + 1)} \sum_{j \in N \setminus M} Y_{ij}, \tag{7}$$

and for all $i \in N \setminus M$,

$$\sum_{j \in M} Y_{ij} < 2w. \tag{8}$$

Profile s where $m > w$ is a SSRE if and only if for all $i \in M$,

$$\sum_{j \in M \setminus \{i\}} Y_{ij} > \frac{2m(m - 1)^2}{wx}. \tag{9}$$

The conditions for a SSRE with $m = w$ are understood as follows. Participant non-deviation (7) requires that the gain from reciprocating the kindness of other participants (who intentionally provide the good) and the material gain from provision, be greater than the loss of not reciprocating the unkindness of non-participants (who intentionally do not reduce the participant’s cost-share). Non-participant non-deviation (8) demands the material savings be greater than the reciprocity cost of not reciprocating participant kindness. As non-participants are not pivotal for each other’s material payoffs ([Lemma 1\(b\)](#)) they provide no reciprocity incentives to one another.

For a SSRE with $m > w$ a non-deviation condition is not needed for non-participants since each non-participant believes no-one thought they were pivotal for his payoff ([Lemma 2\(b\)](#)), implying zero perceived kindness. Participant non-deviation (9) requires the gain from reciprocating the kindness of fellow participants be greater than the material cost.

Note that [Proposition 2](#) does not guarantee existence of SSRE. To see one instance of non-existence suppose that for all $i, j \in N$, $Y_{ij} \geq 2w$, so that (8) does not hold. Then take x sufficiently small such that (9) does not hold.

Cliques coordinating a selfish world

[Examples 1 and 3](#) demonstrated the potential of a reciprocity clique to coordinate behaviour among selfish players. The next two results describe a fairly rich set of circumstances where this is true.

Proposition 3 (*Clique with w members*). *Let $M' \subset N$ be such that $|M'| = w$ and suppose that,*

- (a) for all $i \in N \setminus M'$ and $j \in N$, $Y_{ij} = 0$,
- (b) for all $i \in M'$ and $j \in N \setminus M'$, $Y_{ij} \geq 2w$,
- (c) for all $i \in M'$,

$$\sum_{j \in M' \setminus \{i\}} Y_{ij} > \frac{x(n - w)}{(w + 1)(1 - x)^2} \sum_{j \in N \setminus M'} Y_{ij} - \frac{2}{1 - x}. \tag{10}$$

There then exists a unique SSRE profile s where $M = M'$.

The class of networks referred to in [Proposition 3](#) involves a clique M' of size w with (a) selfish non-clique players, (b) each clique member being sufficiently reciprocal towards each non-clique player, and (c) each clique member being sufficiently reciprocal towards the rest of the clique.

To understand why such networks imply uniqueness see [Example 3](#) (which satisfies conditions (a)–(c) of [Proposition 3](#)) or reason as follows. Condition (a) implies that if there were more than w participants, a participating non-clique player would deviate. Condition (b) implies that if there were w participants and a clique member were not participating he would deviate to reciprocate the kindness of participating non-clique players ((8) is violated). Provision by the clique is thus the unique SSRE, with condition (a) guaranteeing that non-participants do not deviate, and condition (c) ensuring that participants do not either.

It is instructive to study (10) to appreciate what determines the level of intra-clique sensitivity needed to sustain the unique SSRE. A higher n or lower w implies a larger number of unkind non-clique players, increasing the incentive for clique members to deviate. A higher intra-clique sensitivity is thus required to prevent deviation. A higher x implies a decrease in the net material gain from provision, an increase in the unkindness of non-clique players and a decrease in the kindness of clique players. All these effects increase clique members' incentive to deviate, thus once again, a higher clique sensitivity is required to sustain the unique SSRE.

Our next result, [Proposition 4](#), will consider uniqueness with provision by a clique of size greater than w . To better understand the rationale for the class of networks that we will describe in the proposition, it is helpful to return to the networks of [Proposition 3](#) and understand why analogous structures could not give uniqueness with a clique of size greater than w .

Condition (c) of [Proposition 3](#) allows the clique to have homogeneous or heterogeneous intra-clique reciprocity sensitivities. Homogeneity may be inconsistent with uniqueness if the clique is of size greater than w . For the sharpest illustration of why, consider a clique of size $w + 2$ with homogeneous intra-clique sensitivities. Condition (9) shows that the intra-clique sensitivity required for a SSRE with the entire clique participating is greater than that required for one with only $w + 1$ clique members participating. Hence, if there exists a SSRE with the entire clique providing, it is not unique. This issue did not arise with a clique of size w ([Proposition 3](#)) since a proper subset participating is insufficient for provision.

With appropriate heterogeneity, unique provision by a larger clique is possible. This is demonstrated by the special case examined in [Proposition 4](#). Clique members are arranged in a “circle” and each member i is particularly reciprocal towards one of his “neighbours” (called i^+ in [Proposition 4](#)).

Proposition 4 (Clique with $> w$ members). Let $M' = \{1, 2, \dots, m'\}$ where $m' > w$. Define $i^+ = (i \bmod m') + 1$. Suppose that

- (a) for all $i \in N \setminus M'$ and $j \in N$, $Y_{ij} = 0$,
- (b) for all $i \in M'$ and $j \in N \setminus M'$, $Y_{ij} \geq 2w$,
- (c⁻) for all $i \in M'$, $\sum_{j \in M' \setminus \{i, i^+\}} Y_{ij} \leq 2w(w + 1)/x$,
- (c⁺) for all $i \in M'$, $Y_{ii^+} > \max \left\{ 2w, \left(2m'(m' - 1)^2 / wx \right) - \sum_{j \in M' \setminus \{i, i^+\}} Y_{ij} \right\}$.

There then exists a unique SSRE profile s where $M = M'$.

The class of networks referred to in [Proposition 4](#) involves a clique M' of size greater than w with (a) selfish non-clique players, (b) each clique member being sufficiently reciprocal towards each non-clique player, (c⁻) each clique member i being not too reciprocal towards the rest of the clique excluding one of his neighbours, i^+ , (c⁺) each clique member i being sufficiently reciprocal towards one of his neighbours, i^+ .

Conditions (c⁻) and (c⁺) of [Proposition 4](#) ensure clique member i 's sensitivities towards the clique are sufficiently high for a SSRE with the entire clique participating ((9) holds). That clique member i is highly reciprocal towards i^+ plays a key role in avoiding the multiplicity that we discussed with homogeneous intra-clique sensitivities. To see this, consider a profile where only a proper subset of the clique participates. There must then be some participating player i whose neighbour, i^+ , is not participating. Condition (c⁻) then implies that i 's reciprocity towards participants is not sufficiently high to prevent him deviating ((9) violated), thus this profile cannot be a SSRE.

To see that the clique provision SSRE is indeed unique, exclude remaining profiles as follows. Condition (a) implies that provision by more than w participants with at least one non-clique player participating cannot be SSRE, as this player would deviate ((9) violated). Conditions (b) and (c⁺), imply that w participant profiles cannot be SSRE as a non-participating clique member would deviate ((8) violated). There is thus a unique SSRE.

[Propositions 3 and 4](#) demonstrate that for any n , w and x there always exist some reciprocity networks that give uniqueness under cost-sharing. The propositions allow for many constellations of Y_{ij} since most the conditions impose only upper or lower bounds on sensitivities. Note also that the network architectures in the propositions are not necessary for uniqueness. For instance, the network in [Example 4](#) with $w = 3$ gave uniqueness despite it not conforming to the propositions' architectures. Finally, since we consider *strict* SRE, uniqueness should hold for any small perturbation to the weights of the reciprocity network.

Coordination difficulties

Our insights on reciprocity cliques may make coordination seem easy. In general, however, coordination via reciprocity networks is far from trivial. We have already noted equilibrium existence problems. Here follow three negative observations of reciprocity networks where instead the coordination problem is very much alive, beginning with alienation.

Observation 1 (Reciprocity alienation). *If there exist two sets of reciprocity alienated players L' and L'' , such that $L' \neq L''$ and $|L'| = |L''| = w$, then there are multiple SSRE.*

To see the logic behind the statement, simply note that provision by an alienated group of size w is a SSRE and multiplicity follows immediately if there are two such groups (e.g. [Example 2](#)).

Having reciprocity incentives throughout the set of players is clearly important to avoid this outcome. The conditions in [Observation 1](#) are satisfied more easily than it may seem. For example, at least $w + 1$ players with standard preferences towards whom no-one is particularly reciprocal is sufficient. Sufficiently high reciprocity among all players is however not enough to prevent multiplicity as our next observation shows.

Observation 2 (High reciprocity). *If for all $i, j \in N$, $Y_{ij} > 2(w + 1)/x$, then there exists a unique SSRE if and only if $w = n - 1$.*

When $w = n - 1$ candidate SSRE with $m = w$ are broken as by deviating non-participants reduce participants' cost-shares and reciprocate their kindness ([8](#)), leaving only a SSRE with $m = n$, ([7](#)). When $w < n - 1$, ([9](#)) implies that existence of a SSRE with $m = w + 1$ is guaranteed if for all $i, j \in N$, $Y_{ij} > 2(w + 1)/x$, however there are many such equilibria (see e.g. [Example 4](#) with $w = 2$).

A direct implication of [Observation 2](#) is that for a high reciprocity network, provision of a public good which requires almost full participation may be easier than one that requires low participation. Our next observation points out that low participation public goods are also more difficult to provide when preferences are homogeneous.

Observation 3 (Homogeneous reciprocity). *If $w < n - 1$ and for all $i, j \in N$, $Y_{ij} = Y$, then there exist multiple SSRE.*

With homogeneous preferences a unique SSRE cannot have less than n participants otherwise another SSRE would exist where at least one participant and non-participant exchange strategies. Furthermore, full participation cannot be a unique SSRE since the level of reciprocity required for equilibrium is increasing in the number of participants, ([9](#)), so there would also exist equilibria with fewer participants. Preference heterogeneity is thus often key to coordination via reciprocity networks.

4. Extensions

4.1. Welfare

We have so far implicitly defined efficiency in terms of material payoffs. This is not obvious given that utility functions are assumed to be a combination of both material and reciprocity payoffs. While some scholars have argued that welfare should be defined on full utility functions (e.g. [Rabin, 1993, p. 1294](#)), more generally the jury remains out on whether non-material payoffs in social preference models should be included in normative analysis ([Bernheim and Rangel, 2005](#)).

If one were to define efficiency on full utility functions, it is important to recognise that utilities are a function of kindness and perceived kindness, which themselves are a function of both beliefs and strategies. Any efficiency definition must thus account for beliefs as well as strategies. In this section we shall consider the implications of Pareto efficiency defined on full utility functions where beliefs are correct.

Using this revised efficiency definition we can illustrate an interesting implication of including reciprocity payoffs. Recall P&R with neither refunds nor cost-sharing. Let $n = 4$, $w = 2$, $x = 0.5$, player 1 be alienated and the others form a clique of strength 3. Player 1 only participating, so that players 2, 3 and 4 are all unkind to each other, is now efficient as one can show that at least one clique member must get strictly lower utility in all other profiles. Reciprocation of unkindness may not seem “normatively desirable.”

In the remainder of this section we explain the main implications of our reciprocity-inclusive efficiency definition for our game. With neither refunds nor cost-sharing it remains the case that any w participant SRE is Pareto superior to the no-participation SRE. To see this, note that all players receive strictly higher material payoffs in former profile; non-participants receive zero reciprocity payoff in both profiles; and participants reciprocate each others' kindness so receive a higher reciprocity payoff in former profile.

Although the inefficiency problem remains, P&R's refunds may no longer recover efficiency. For example, take P&R with refunds but not cost-sharing, let $n = 3$, $w = 2$, $x = 0.5$ and all $Y_{ij} = 0$ except $Y_{21} > 8$. Despite players 1 and 2 only participating and players 1 and 3 only participating both being SSRE, the former now Pareto dominates the latter as 1 receives the same utility and 2 and 3 receive higher utility (2 a higher reciprocity payoff and 3 a higher material payoff).

While the set of SSRE may be (partially) Pareto rankable, transitivity of the Pareto criterion implies that there exists some subset of SSRE undominated by other SSRE. This subset is in fact efficient overall. Take an SSRE from the subset and rule out Pareto improvements as follows: if $m > w$, a non-participant in the SSRE who now participates gets strictly lower material payoff and an identical reciprocity payoff ([Lemmas 1\(b\) and 1\(c\)](#)); if $m < w - 1$, a participant in the SSRE gets strictly lower material payoff and no higher reciprocity payoff ([Lemma 1\(a\)](#)); if $m = w - 1$, a participant in the SSRE who is now a non-participant gets a strictly lower material payoff and a no higher reciprocity payoff ([Lemma 1\(a\)](#)).

In Section 3 the set of efficient profiles was fixed and the reciprocity network determined equilibrium behaviour. With reciprocity included in the efficiency definition, the reciprocity network simultaneously determines which strategy profiles are efficient and equilibrium behaviour. Is a unique and efficient SSRE still attainable with cost-sharing? Recall Example 3 from Section 3.3 to see that the answer is yes: $n = 4$, $w = 2$, $x = 0.5$ and for all $i, j \in N$, $Y_{ij} = 0$ except $Y_{13} = Y_{24} = 5$ and $Y_{12} = Y_{21} = 6$, implying a unique SSRE with the clique participating. This is efficient since any other profile either involves non-provision or player 3 and/or 4 participating, hence 3 and/or 4 being strictly worse off. Since Proposition 3 is established by generalising the logic of this example, a unique and efficient SSRE remains possible.

4.2. Networks of social preferences

One might be curious to know whether our results with reciprocity are similar to those under other social preferences models. As one may expect, our findings do not generally hold for other models. We identify some similarities and differences to better understand what drives our results on reciprocity.

Consider simple forms of altruism (cf. Andreoni, 1989) and difference-aversion (cf. Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000). An altruistic player i 's preferences are represented by $U_i(s) = \pi_i(s) + \sum_{j \neq i} \alpha_{ij} \pi_j(s)$, where $\alpha_{ij} \geq 0$ denotes i 's altruism towards player j . A difference-averse i 's preferences are represented by $U_i(s) = \pi_i(s) - \sum_{j \neq i} \gamma_{ij} (|\pi_i(s) - \pi_j(s)|)$, where $\gamma_{ij} \geq 0$ is i 's sensitivity to payoff differences with j .

Given that our objective is to identify the drivers of our results, we use forms of altruism and difference-aversion that seem comparable to D&K in the following senses: material and social preference payoffs are additively separable; i is permitted to have heterogeneous preference parameters towards different j ; and the domain of preference parameters is not bounded from above.⁷

One might conjecture that equilibria with these distributional social preference models will always differ from reciprocity since some D&K incentives cannot exist in such models (e.g. Example 0). For a case where the sets of equilibria are identical, consider altruistic players with refunds only. Deviation from an efficient profile, those with w participants, reduces the deviator's material payoff and cannot increase others' payoffs, thus the profile is a strict equilibrium. As no other profiles are strict equilibria, the set of strict equilibria is independent of α_{ij} (cf. Proposition 1(ii)), hence P&R's refund result is robust to altruism.

For a very different outcome, consider difference-averse players with refunds only. Let $n = 3$, $w = 2$, $x = 0.5$ and for all $i, j \in N$, $\gamma_{ij} > 1$. There is then a unique inefficient strict equilibrium with full participation! Materially efficient profiles are not equilibria as deviation secures payoff equality.

Comparing the above, we see that the role of pivotality over others' material payoffs in D&K is a key driver of Proposition 1. For altruism and reciprocity, i 's strategy cannot influence his social preference payoff if he is non-pivotal. Non-participants in a w participant profile thus have neither reciprocity (Lemmas 1(b) and 1(c)) nor altruism incentives to deviate. For models where pivotality is not necessary for influencing the social preference payoff (e.g. difference-aversion), non-participants may deviate and P&R may not be robust.

Altruism and reciprocity are not similar under cost-sharing. For instance, consider high social preferences to see that both altruism and difference-aversion produce different result to D&K. Recall Example 4 ($n = 4$, $w = 2$ and for all i, j , $Y_{ij} = 13$) and contrast the predictions to those of altruism and difference-aversion with sufficiently high α_{ij} and γ_{ij} . All profiles with 3 participants are SSRE as the non-participant believes that no participant thought himself pivotal for the non-participants' payoff (Lemma 2(b)). By contrast, with sufficiently high altruism or difference-aversion, the non-participant deviates to increase others' payoffs and attain payoff equality. There is thus a unique strict equilibrium of full participation. The difference between D&K and other models is driven by the important role of perceptions of others' behaviour in reciprocity theory, a component that is absent in other models.

Now consider reciprocity networks which imply a unique SSRE. Our earliest instance was Example 1 ($n = 3$, $w = 2$, $Y_{12} = Y_{21} = 6$, for all other $i, j \in N$, $Y_{ij} = 0$). Multiplicity was overcome by clique members deviating from efficient profiles unless they play the same strategy, giving a unique SSRE with only the clique participating. A sufficiently strong altruism or difference-aversion clique would analogously imply uniqueness. The models give similar results because cost-sharing implies non-participants are pivotal for participants' payoffs.

For reciprocity we found that a clique was not sufficient for uniqueness due to alienation (Example 2), but having positive reciprocity sensitivities towards non-clique members achieves uniqueness (Example 3, Propositions 3 and 4). Revisit Example 3 ($n = 4$, $w = 2$, $x = 0.5$) to see that analogous networks give uniqueness for altruism but not for difference-aversion. Consider an altruism network where $\alpha_{12}, \alpha_{13}, \alpha_{21}, \alpha_{24} > 2$ and all other $\alpha_{ij} = 0$. In any profile with provision other than only the clique (players 1 and 2) participating, a non-participant clique member or a non-clique player would deviate, we thus have a unique strict equilibrium with the clique providing. To see the network architecture in Example 3 will not give uniqueness with difference aversion, note that player 1 deviating from a profile where only 3 and 4 participate ($\gamma_{13} > 2(1 + \gamma_{12})$) is inconsistent with player 1 not deviating from a profile where only 1 and 2 participate ($\gamma_{13} < 1$). Pivotality again drives the difference. If only 3 and 4 participate, 1 is non-pivotal for 2's payoff so α_{12} and Y_{12} have no

⁷ Unboundedness from above also implies that our conclusions would remain unaffected were we to replace α_{ij} and γ_{ij} by α_{ij}/n and γ_{ij}/n , as is often done in social preference models.

influence on i 's payoff, whereas γ_{12} does. In cases like this example, the additional requirements on parameters imposed by difference-aversion preclude uniqueness.

The above comparisons highlight the important roles played by pivotality and perceptions of others' behaviour for our results on reciprocity and uniqueness. While we can find some instances of agreement with our model, one cannot expect our insights on reciprocity to carry over to very different motivations like altruism and difference-aversion with a high degree of generality.

5. Conclusion

Societies around the world have faced and continue to face great challenges. A fundamental prerequisite to overcome many of these is the coordination of social participation. Palfrey and Rosenthal (1984) (P&R) provided an adequate and specific framework for exploring related issues: discrete-level public goods games, which capture the key problems of private provision of public goods in condensed form.

Reciprocity is an important form of human motivation which many scholars have suggested may matter in public goods provision settings. We have explored how the application of a formal model of reciprocity (Dufwenberg and Kirchsteiger, 2004; cf. Rabin, 1993) affects predictions in P&R's context, with extensions.

A new potential source of coordination power was uncovered: the interaction of networks of reciprocity incentives and cost-sharing. In the absence of cost-sharing, reciprocity does nothing to solve the coordination problem, P&R is robust. In the absence of reciprocity, cost-sharing has no effect on coordination. But together, behaviour can be coordinated.

Since cost-sharing can potentially be manipulated by institution designers, understanding its effects seems particularly important. While there is some experimental work in continuous contribution games (Marks and Croson, 1998; Spencer et al., 2009), the conditions under which cost-sharing can coordinate behaviour are poorly understood. Our work suggests that empirical exploration of its interaction with reciprocity may be worthwhile.

The precise nature of the reciprocity network matters for coordination. Properties of the network can both help coordination and hinder it. We are then left with a pressing empirical question: what does the reciprocity network look like in reality? Given its importance and the absence of empirical research on its nature, addressing this may be of interest for public good scholars.

Societies probably have new challenges waiting for them. We hope that by understanding how social participation can be coordinated, resolutions are less of a struggle, leaving more time to enjoy the public goods produced.

Appendix A

Proof of Proposition 1 (Reciprocity in P&R). (i) *No refunds.* We identify the sets of SRE and SSRE using arguments that are independent of Y_{ij} . First consider a candidate equilibrium strategy profile with no participants. Deviation by i gives him a strictly lower material payoff given that $w > 1$ and does not affect his reciprocity payoff since for all j , $\kappa_{ij}(0, s_{-i}) = \kappa_{ij}(1, s_{-i}) = 0$ (Lemma 1(a)). Thus non-participation is a SRE as well as a SSRE. Second, consider a candidate equilibrium with w participants. Deviation by i gives him a strictly lower material payoff. If i deviates from non-participation, his reciprocity payoff is unchanged since for all j , $\kappa_{ij}(0, s_{-i}) = \kappa_{ij}(1, s_{-i}) = 0$ (Lemmas 1(b)+1(c)). If i deviates from participation, his reciprocity payoff is strictly lower since for all j such that $s_j = 1$, $\lambda_{iji}(1, s_{-j}) > 0$ and $\kappa_{ij}(1, s_{-i}) > \kappa_{ij}(0, s_{-i})$, and for all j such that $s_j = 0$, $\lambda_{iji}(0, s_{-j}) = 0$ (Lemma 2(c)). Thus w participants is a SRE and a SSRE. To see that there are no more equilibria, take a candidate equilibrium with some number of participants other than 0 or w . If i deviates from participation his material payoff is strictly higher and his reciprocity payoff is unchanged given that for all j , $\kappa_{ij}(0, s_{-i}) = \kappa_{ij}(1, s_{-i}) = 0$ (Lemmas 1(a)-1(c)). Thus this profile is neither SRE nor SSRE.

(ii) *Refunds.* We identify the sets SSRE using arguments that are independent of Y_{ij} . Consider a candidate SSRE profile with w participants. Deviation by i gives him a strictly lower material payoff and no increase in reciprocity payoff. More specifically, if i deviates from non-participation his reciprocity payoff is unchanged since for all j , $\kappa_{ij}(0, s_{-i}) = \kappa_{ij}(1, s_{-i}) = 0$ (Lemmas 1(b)+1(c)); and, if i deviates from participation his reciprocity payoff is strictly lower since for j such that $s_j = 1$, $\lambda_{iji}(1, s_{-j}) > 0$ and $\kappa_{ij}(1, s_{-i}) > \kappa_{ij}(0, s_{-i})$, and for j such that $s_j = 0$, $\lambda_{iji}(0, s_{-j}) = 0$ (Lemma 2(c)). Thus w participants is a SSRE. To see that there are no more SSRE, first consider a candidate SSRE with more than w participants. If i deviates from participation, his material payoff increases and his reciprocity payoff is unchanged since for all j , $\kappa_{ij}(0, s_{-i}) = \kappa_{ij}(1, s_{-i}) = 0$ (Lemmas 1(b)+1(c)). Thus this profile is not a SSRE. Second, consider a candidate SSRE with a number of participants strictly between 0 and w . If i deviates from participation both his material and reciprocity payoffs are unchanged since for all j , $\kappa_{ij}(0, s_{-i}) = \kappa_{ij}(1, s_{-i}) = 0$ (Lemma 1(a)). Thus this profile is not a SSRE. Finally, consider the candidate SSRE with no participants. Deviation by i gives him identical material and reciprocity payoffs since for all j , $\kappa_{ij}(0, s_{-i}) = \kappa_{ij}(1, s_{-i}) = 0$ (Lemma 1(a)). Thus this profile is not a SSRE. □

Proof of Proposition 2 (Cost-sharing equilibria). First show that non-provision profiles cannot be SSRE. Consider a candidate SSRE with no participants. Deviation by i gives him the same material and reciprocity payoffs since for all j , $\kappa_{ij}(0, s_{-i}) = \kappa_{ij}(1, s_{-i}) = 0$ (Lemma 1(a)). Thus this is not a SSRE. Consider a candidate SSRE with a number of participants strictly between 0 and w . If i deviates from participation both his material and reciprocity payoffs are unchanged since for all j , $\kappa_{ij}(0, s_{-i}) = \kappa_{ij}(1, s_{-i}) = 0$ (Lemma 1(a)). Thus this is not a SSRE.

Then identify conditions for provision profiles to be SSRE. A SSRE profile with w participants requires that for all i such that $s_i = 1$, $u_i((1, s_{-i}), (s_{-i}, (s_{-j})_{j \neq i})) > u_i((0, s_{-i}), (s_{-i}, (s_{-j})_{j \neq i}))$, hence (7), and that for all i such that $s_i = 0$, $u_i((0, s_{-i}), (s_{-i}, (s_{-j})_{j \neq i})) > u_i((1, s_{-i}), (s_{-i}, (s_{-j})_{j \neq i}))$, hence (8). In a SSRE profile with strictly more than w participants, if i deviates from non-participation his material payoff is strictly lower and his reciprocity payoff is unchanged since for all j , $\lambda_{iji}(0, s_{-j}) = \lambda_{iji}(1, s_{-j}) = 0$ (Lemma 2(b)). Thus a SSRE requires only that for all i such that $s_i = 1$, $u_i((1, s_{-i}), (s_{-i}, (s_{-j})_{j \neq i})) > u_i((0, s_{-i}), (s_{-i}, (s_{-j})_{j \neq i}))$, hence (9). \square

Proof of Proposition 3 (Clique with w members). Reason as follows to see that there is a unique SSRE. First consider a candidate SSRE profile with more than w participants. There must exist some $i \in N \setminus M'$ for whom $s_i = 1$. But then for such i , (a) implies that the LHS of the inequality in (9) is equal to zero. Thus this is not a SSRE. Second, consider a candidate SSRE profile with w participants where the set of participants is not equal to M' . There must exist some $i \in M'$ for whom $s_i = 0$ and some $j \in N \setminus M'$ for whom $s_j = 1$. But then for such i , (b) implies that the LHS of the inequality in (8) is at least $2w$. Thus this is not a SSRE. Consider the only remaining candidate SSRE where the set of participants equals M' . For all $i \in M'$ inequality (7) is satisfied given (b) and (c). The inequality in (8) is satisfied given (a). Thus this is the unique SSRE. \square

Proof of Proposition 4 (Clique with $> w$ members). Reason as follows to see that there is a unique SSRE. First consider a candidate SSRE where the set of participants is a proper subset of M' with strictly more than w participants. Given (c⁻) there must be some $i \in M'$ such that $s_i = 1$ for whom $\sum_{j \in M \setminus \{i\}} Y_{ij} \leq 2w(w+1)/x$. But then for such i , inequality (9) does not hold. Thus this is not a SSRE. Second, consider a candidate SSRE where there are strictly more than w participants and the set of participants is not a subset of M' . There must be some $i \in N \setminus M'$ such that $s_i = 1$. But then for such i , inequality (9) does not hold given (a). Thus this is not a SSRE. Third, consider a candidate SSRE where there are w participants. There must exist some $i \in M'$ such that $s_i = 0$ for whom inequality (8) does not hold given (b) and (c⁺). Thus this is not a SSRE. The final candidate SSRE profile is where the set of participants is M' . Inequality (9) holds given (c⁻) and (c⁺). This is the unique SSRE. \square

References

- Andreoni, J., 1989. Giving with impure altruism: applications to charity and Ricardian equivalence. *J. Polit. Economy* 97, 1447–1458.
- Bartling, B., Fischbacher, U., Schudy, S., 2015. Pivotality and responsibility attribution in sequential voting. *J. Public Econ.* 128, 133–139.
- Battigalli, P., Dufwenberg, M., 2009. Dynamic psychological games. *J. Econ. Theory* 144, 1–35.
- Bernheim, D.B., Rangel, A., 2005. Behavioral public economics: welfare and policy analysis with non-standard decision makers. NBER Working Paper Series, Working Paper 11518.
- Bolton, G.E., Ockenfels, A., 2000. ERC: a theory of equity, reciprocity and competition. *Amer. Econ. Rev.* 90, 166–193.
- Bourlès, B., Bramoullé, Y., Perez-Richet, E., 2015. Altruism in networks. Mimeo.
- Dufwenberg, M., Kirchsteiger, G., 2000. Reciprocity and wage undercutting. *Europ. Econ. Rev.* 44, 1069–1078.
- Dufwenberg, M., Kirchsteiger, G., 2004. A theory of sequential reciprocity. *Games Econ. Behav.* 47, 268–298.
- Falk, A., Fischbacher, U., 2006. A theory of reciprocity. *Games Econ. Behav.* 54, 293–315.
- Fehr, E., Schmidt, K.M., 1999. A theory of fairness, competition and cooperation. *Quart. J. Econ.* 114, 817–868.
- Fischbacher, U., Fehr, E., Gächter, S., 2001. Are people conditionally cooperative? Evidence from a public goods experiment. *Econ. Letters* 71, 397–404.
- Fischbacher, U., Gächter, S., 2010. Social preferences, beliefs, and the dynamics of free-riding in public goods experiments. *Amer. Econ. Rev.* 100, 541–556.
- Geanakoplos, J., Pearce, D., Stacchetti, E., 1989. Psychological games and sequential rationality. *Games Econ. Behav.* 1, 60–79.
- Keser, C., van Winden, F., 2000. Conditional cooperation and voluntary contributions to public goods. *Scand. J. Econ.* 102, 23–39.
- Ledyard, J., 1995. Public goods: a survey of experimental research. In: Kagel, J., Roth, A. (Eds.), *Handbook of Experimental Economics*. Princeton University Press.
- Leider, S., Möbius, M.M., Rosenblat, T., Do, Q.-A., 2009. Directed altruism and enforced reciprocity in social networks. *Quart. J. Econ.* 124, 1815–1851.
- Makris, M., 2009. Private provision of discrete public goods. *Games Econ. Behav.* 67, 292–299.
- Marks, M., Croson, R., 1998. Alternative rebate rules in the provision of a threshold public good: an experimental investigation. *J. Public Econ.* 67, 195–220.
- Palfrey, T., Rosenthal, H., 1984. Participation and the provision of discrete public goods: a strategic analysis. *J. Public Econ.* 24, 171–193.
- Rabin, M., 1993. Incorporating fairness into game theory and economics. *Amer. Econ. Rev.* 83, 1281–1302.
- Rothenhäusler, D., Schweizer, N., Szech, N., 2013. Institutions, shared guilt and moral transgression. WZB discussion paper, no. SP II 2013-305.
- Shinohara, R., 2009. The possibility of efficient provision of a public good in voluntary participation games. *Soc. Choice Welfare* 32, 367–387.
- Sobel, J., 2005. Interdependent preferences and reciprocity. *J. Econ. Lit.* 63, 392–436.
- Spencer, M.A., Swallow, S.K., Shogren, J.F., List, J.A., 2009. Rebate rules in threshold public good provision. *J. Public Econ.* 93, 798–806.
- Sugden, R., 1984. Reciprocity: the supply of public goods through voluntary contributions. *Econ. J.* 94, 772–787.