Cooperative Networks: Theory and Experimental Evidence*

Katinka Pantz† Anthony Ziegelmeyer‡

Very preliminary draft. Comments are welcome.

Abstract

We consider a modified pure public good game characterized by a pre-play negotiation stage, on which *pairs of players* can form binding cooperation commitments. As the introduced mechanism only supports pairwise rather than more inclusive commitments, it does not implement the efficient outcome. We theoretically derive the incentive compatible and efficient *cooperative networks* and evaluate the behavioral efficacy of the suggested mechanism to promote and stabilize cooperation. We present the results of two experiments. The first implemented environment follows the standard procedures for the voluntary contributions mechanism and establishes that neither repetition with an unknown end nor voluntary costly monitoring are behaviorally sufficient to induce cooperative outcomes. In the second experiment we introduce the pairwise commitment mechanism and we observe aggregate cooperation rates beyond those observed in the first experiment whatever the monitoring cost but also beyond the rates supported by the formation of incentive compatible networks when individual monitoring is relatively cheap. Groups differ largely in their ability to make use of the pairwise commitment mechanism: while some groups converge to full cooperation by managing to coordinate on the formation of efficient networks over time, both networks and cooperation rates unravel in other groups. An extended version of our theoretical setting with inequity averse players in the form suggested by Fehr and Schmidt (1999) captures the stylized facts of both experiments.

Keywords: Strategic formation of networks; Social dilemma; Positive externalities; Experiments.

JEL Classification: C92; D85; H41; Z13.

*This research was undertaken while the first author was also affiliated with the Max Planck Institute of Economics whose funding is gratefully acknowledged. Help in conducting the experiments has been provided by Bettina Bartels, Frederic Bertels, Håkan Fink and Evelin Wacker. Stephane Bertrand has been part of this project at an early stage. We are very grateful for his contribution and the nice discussions. In addition, we would like to thank Bradley Ruffle, participants of the 2005 ESA Meeting in Montreal, of the 2005 SAET conference in Vigo and seminar participants at MPI Jena and Humboldt University Berlin for their comments and suggestions. The usual disclaimer applies.

†Institute of Public Economics, Department of Economics, Humboldt University of Berlin, Spandauer Strasse 1, 10178 Berlin. Email: kpantz@web.de.

‡Max Planck Institute of Economics, Strategic Interaction Group, Jena, Germany. Email: ziegelmeyer@econ.mpg.de.

1 Introduction

Individual contributions to joint endeavors are exposed to potential free-riding in a broad range of contexts. One solution is for the state to intervene and, e.g., provide public goods. But this is neither generally defensible - as some public goods serve too highly localized interests - nor desirable. Extensive field work has shown that local communities frequently develop elaborate organizational patterns in relation to the specific problems and opportunities posed by the governance of various common pool resources or the provision of public goods. These organizational patterns are informal, if one considers only state instruments as formal, but reflect enduring and in some cases highly efficient forms of organization. In fact, recent empirical evidence suggests that state intervention may frustrate rather than facilitate the private provision of public goods, especially where it ignores existing forms of scarce institutional capital (Montgomery and Bean, 1999; Ostrom, 2000). Given that informally organized self-governance - even though widespread - is far from inevitable, empirical research has focused on identifying those variables and social mechanisms that are decisive for its establishment and maintenance. One social mechanism which clearly fosters cooperation in the field is the ability of individuals to negotiate and form cooperation agreements. This paper studies - both theoretically and experimentally - the efficacy of a mechanism which allows *individuals to form pairwise cooperation commitments*, those cooperation commitments being beneficial for the entire community.

Coase (1960) has early on put forward the equally simple and influential idea that collective action problems can be resolved when the involved individuals are given the possibility to negotiate.[1] But his solution remained incomplete as he did not specify the negotiation mechanism and the type of agreements that can be reached. A recent strand in the theoretical literature attempts to fill some of the gaps. Dixit and Olson (2000) concentrate on the crucial but often insufficiently acknowledged fact that any negotiation and therewith the formation of cooperation agreements is *voluntary.* The authors study a two-stage model of a discrete public good with N identical individuals, and they assume that there is an integer $M < N$ such that a coalition of size M finds it profitable to provide the public good, sharing the costs equally. The first stage is non-cooperative, where isolated individuals decide whether to participate in the second stage. If M or more individuals have chosen to participate at the first stage, it is optimal for them to produce the good in the second stage. If fewer than M individuals choose to participate, the good is not produced. Dixit and Olson show that as N increases and $M < N$, the likelihood that the good gets provided goes down very rapidly and an efficient outcome is extremely unlikely. Ray and Vohra (2001) study the provision of public goods without a provision point in a much richer model of coalition formation. Individuals are allowed to not only form one but several coalitions, where the interaction between coalitions is assumed to be non-cooperative, while the members of a given coalition commit to maximize their coalition's joint interest. Only in some cases does the grand coalition form in equilibrium, resulting in efficient provision of the public good. In other cases, the equilibrium consists of several coalitions and inefficient provision. Taken together, the two theoretical studies show that full efficiency is not always obtained when individuals have the discretion over whether or not to enter negotiations in the first place.

[1] Coase assumes that property rights are well-defined and negotiation is costless.

Dixit and Olson's (2000) as well as Ray and Vohra's (2001) theoretical findings replicate a stylized empirical fact: village-wide welfare and insurance arrangements are largely absent in the field (Fafchamps, 1992; Murgai, Winters, Sadoulet, and de Janvry, 2002; Fafchamps and Lund, 2003). Still, the two models' assumption according to which the organizational mode of successful cooperative groups is based on multilateral agreements is, in many cases, rejected by field data. Indeed, a large fraction of the cooperative projects are constituted by sets of pairwise cooperation agreements. An illustrative example are South Indian communities which successfully solve the problem of externalities inherent in mixed lifestock and arable husbandry where holdings are scattered, fields unfenced and where fodder is short (Wade, 1988).[2] While some of the communities that successfully employ mixed husbandry have what Wade calls a 'corporate structure' characterized by field guards deployed on assembly decision and paid by village funds, many communities reduce interdependencies by mutual pairwise restraint between neighbors which proceeds from "the danger that A will damage B's crops if B allows his lifestock to damage A's".[3] In the latter case, cooperation does not result from explicit arrangements across several people (as reached in assemblies) but cooperating individuals are connected to a small number of other people who, in turn, are themselves connected to still other people. Connections are given by direct and hence highly personalized cooperation agreements where each mutual pairwise agreement is part of a complex set of comparable agreements which together form a *cooperative network*, which only as a whole guarantees the effectiveness of cooperation. In conclusion, the empirical evidence suggests that a coalition-based response to interdependencies with the cooperating group acting as a single unit is neither self-evident, nor the most frequently observed organizational mode. And the proposed explanation is obvious: if a group contains diverse preferences with respect to, e.g., the desirable provision level of a public good, reaching a consensus becomes increasingly difficult the more individuals there are whose preferences must converge.[4]

In this paper, we suggest a theoretical framework in the tradition of Dixit and Olson (2000) and Ray and Vohra (2001). We consider a 2-stage game in which individuals can non-cooperatively decide to form *bilateral cooperation agreements* on a first stage. By forming a cooperation agreement the involved pair of individuals agrees to set their respective subsequent contribution levels to a pure public good to maximize the joint payoff of the pair. We assume that agreements are binding and hence have the nature of cooperation *commitments*.[5] In our setting, the pairwise commitment

[2]Scattering, i.e., the division of each farmer's holding into several plots which are spread around the village is common in peasant societies. The underlying rationale is to diversify the individual farmer's risk of crop loss. A direct implication is that fencing becomes prohibitively expensive.

[3]Next to animal herding in mixed husbandry also irrigation canal-cleaning activities of farmers are typically resulting from agreements with one or two neighbors in which each will clean the canal adjacent to his or her own land were all pairwise agreements taken together guarantee the functioning of the irrigation system (Wade, 1988; Ostrom, 1990; Tang, 1992).

[4]Instead, pairwise cooperation agreements can support 'partial cooperation', accommodating the fact that some individuals may be willing to cooperate in spite of a large number of free-riders while others may require a large fraction of co-users to cooperate as a precondition for their own cooperation. Already David Hume argued in these lines when claiming that "two neighbours may agree to drain a meadow, which they possess in common; because 'tis easy for them to know each others mind; and each must perceive, that the immediate consequences of his failing in his part, is the abandoning of the whole project. But 'tis very difficult, and indeed impossible, that a thousand persons shou'd agree in any such action; it being difficult for them to concert so complicated a design, and still more difficult for them to execute it; while each seeks a pretext to free himself of the trouble and expense, and wou'd lay the whole burden on others." (Hume, 1928)

[5]In line with Dixit and Olson's and Ray and Vohra's models, we assume rather then deduce that agreements are binding and efficient. In this first approach we refrain from explicitly modeling the mechanism that is inducing

mechanism renders the free-rider problem solvable at the pair but not at the population level and consequently does not implement the efficient outcome.

Qualifying our theoretical results we experimentally evaluate the *actual behavioral efficacy of the suggested pairwise commitment mechanism in inducing and stabilizing cooperation.* To address this research question, we report the results of two separate experiments. The objective of the first experiment is to provide all necessary prerequisites, viz., to establish a standard of comparison, to identify potentially relevant explanatory approaches and to settle some methodological issues. We implement a repeated non-linear pure public good game with an unknown end under the standard voluntary contribution mechanism, in which individuals have no possibility to form pairwise commitments, and we manipulate the availability of information about others' behavior. The objective is to test the role of 1) voluntary but costly individualized monitoring and of 2) repetition with an unknown end as independent mechanisms for solving the public good provision problem. Both features serve to more closely reflect likely effects in naturally occurring social dilemma situations, where behavioral information is - unlike in most voluntary public good provision experiments - not automatically and costlessly available and where individuals have no ex-ante information about the actual length of their interaction. Our results of the first experiment show that 1) the cost of voluntary monitoring has no significant impact on behavior, and 2) repetition with an unknown end does not induce strategic supergame play, while it helps to circumvent otherwise observed artificial end-game effects. In fact, our findings clearly demonstrate that the observed over-contribution relative to the predicted level of the stage game under the assumption of individual greed (static Nash equilibrium) results from the interplay of self-interested subjects and subjects who are also concerned about the payoffs of others.

In our second experiment, we introduce the possibility to form pairwise cooperation commitments. Our implementation is once again intended to capture essential features of natural negotiation processes. The formation of pairwise cooperation commitments is 1) *optional* meaning that the negotiation stage is entered only on demand rather than mandatorily, and 2) cooperative networks are *renegotiable* in regular intervals. On the (re)negotiation stage we implement a sequential link-based network formation protocol closely related to the seminal protocol suggested by Aumann and Myerson (1988) and adapted by Watts (2002). Accordingly, commitments are 3) *consensual* meaning that a commitment is formed and becomes effective (creates a binding obligation) only if both involved individuals agree on its formation, where the implemented protocol allows experimental subjects to correct errors and to iteratively respond to others' decisions. We find that experimental subjects make use of the commitment mechanism. Furthermore, a considerable fraction of the observed networks are pairwise stable. The implemented mechanism induces and stabilizes aggregate cooperation rates both beyond those observed under the voluntary contribution mechanism and beyond the rate induced by the formation of pairwise stable networks. Subjects especially in pairwise stable networks go beyond the committed level, which suggests that the mechanism provides a social context that supports 'genuine cooperation' beyond what can be accounted for

agreements that are efficient at the pair level. One possibility to do so could be to draw on an appropriate adaptation of the compensation mechanism suggested by Varian (1994). There are many ways to enforce the bindingness of agreements such as, e.g., through contract law but also through posting a bond with a third party, through decentralized coercion from interpersonal pressure or - under some cultural circumstances - even through ethical obligation alone.

by self-interest. The efficacy of the mechanism in this respect depends on the cost of individual monitoring. Where individual monitoring is relatively cheap, the pairwise commitment mechanism *increases and sustains* cooperation at a very high level. When monitoring is comparatively expensive, the mechanism cannot sustain high cooperation rates. We observe a large heterogeneity in the ability of different groups to make use of the pairwise commitment mechanism. While some groups succeed in using the mechanism to coordinate on full cooperation by building efficient networks, networks unravel and cooperation breaks down in other groups.

Finally, in an attempt to capture the stylized facts observed in the laboratory, we go back to our theoretical framework and extend it by incorporating social preferences. Inequity averse individuals in the form suggested by Fehr and Schmidt (1999) contribute to the provision of a pure public good both under the voluntary contribution mechanism and the pairwise commitment mechanism. This extended theoretical setting leads to predictions which are in line with our experimental evidence: 1) cooperation both below and above the static Nash equilibrium can be achieved in the voluntary contribution mechanism, 2) the cost of monitoring only matters in the pairwise commitment mechanism, and 3) with low monitoring costs, cooperative networks can induce full efficiency.

The paper is organized as follows. Section 2 introduces the theoretical setting and derives some general results for the non-linear case that we are considering in our experimental implementation. Sections 3-4 respectively discuss the two experiments. In section 5, we provide an extension of our theoretical setting by incorporating social preferences. We conclude in section 6. Proofs are provided in the Appendix.

2 Theoretical background

In this section, we introduce a game of voluntary efforts which is formally equivalent to a non-linear (pure) public goods provision game. First, we derive the predictions of the game when players individually choose their effort levels. Second, we allow players to form pairwise cooperation commitments regarding their effort levels. Any pair of players which has formed a cooperation commitment jointly sets its effort level to maximize the pair's payoff. The collection of pairwise cooperation commitments defines a cooperative network. We characterize the strategically stable cooperative networks and we analyze the extent to which individual incentives to form cooperative networks align with social efficiency.[6]

2.1 A game of voluntary efforts

Suppose that there are $n > 1$ identical players who simultaneously exert an effort whose benefits accrue equally to all players. We denote by e_i player i's effort level, and we assume that the effort exerted involves a cost of $c(e_i)$ which is private to player $i \in N = \{1, \ldots, n\}$. The effort cost function is assumed to be strictly increasing and strictly convex, and $e_i \in [0, \bar{e}]$ with $c'(0) \leq 1$ and $c'(\bar{e}) \geq n$. Player i's material payoff is given by $\pi_i(e_i, e_{-i}) = \sum_{i \in N} e_i - c(e_i)$ where $e_{-i} = (e_j)_{j \neq i}$ with $j \in N$.

[6] Our theoretical setting borrows heavily from Ray and Vohra (2001) except that we consider pairwise instead of multilateral binding agreements.

2.2 The solution without pairwise cooperation commitments

Each player has a dominant effort level $e^* = \arg\ \max_e e - c(e)$ which according to the first order condition entails $c'(e^*) = 1$ (the second order condition is fulfilled as c is strictly convex).

This dominant effort level can be contrasted with the fully efficient effort level the players would like to commit to if they could, namely the level which maximizes the sum of their material payoffs. Since c is strictly convex, efficiency implies that each player exerts the same effort level in order to maximize the sum of the material payoffs. Each player's fully efficient effort level, which we denote by e^{**}, is the solution of

$$\max_e n\, e - c(e), \tag{1}$$

which according to the first order condition leads to $c'(e^{**}) = n$. Given the assumed shape of the effort cost function, $e^{**} > e^*$ and $\pi_i(\mathbf{e}^{**}) > \pi_i(\mathbf{e}^*)$ where $\mathbf{e}^*$ is the vector of dominant effort levels and $\mathbf{e}^{**}$ is the vector of fully efficient effort levels.

2.3 The solution with pairwise cooperation commitments

In what follows, we depart from the (individualistic) voluntary effort game by permitting each player to make an offer to write a *binding* cooperation agreement with any other player. If player i makes an offer to player j and player j accepts this offer, then, based on their mutual consent, a *cooperation commitment* is formed between the two players. Any pair of players which has formed a cooperation commitment jointly sets its effort level to maximize the pair's material payoff.

Due to the assumed linearity of external effects, the optimal effort level of a pair of players does not depend on what other pairs of players are doing; in terms of the amount of effort to be exerted, each pair has a dominant strategy. In effect, the problem facing a pair of players is to exert an effort level of e_{BA} per player, where e_{BA} solves equation (1) for $n = 2$, i.e., it solves

$$\max_e 2\, e - c(e),$$

which according to the first order condition leads to $c'(e_{BA}) = 2$ where $e^{**} \geq e_{BA} > e^*$.

At the end of the negotiations phase, a *cooperative network* consisting of all pairwise cooperation commitments that have been formed among the players results. A network is a list of *unordered* pairs of players $\{i, j\}$, where $\{i, j\} \in g$ indicates that player i and player j have formed a cooperation commitment, $i, j \in N$ and $i \neq j$. For simplicity, we will from now on denote the cooperation commitment $\{i, j\}$ by ij. $G = \{g \mid g \subseteq g^N\}$ represents the set of all possible cooperative networks on N where g^N denotes the complete network, i.e., the set of all subsets of N of size 2. For each $g \in G$, let $n(g) =\mid \{i \mid \exists j \text{ s.t. } ij \in g\} \mid$ be the number of players who are part of at least one cooperation commitment in g, let $l_i(g) =\mid \{ij \mid \exists j \text{ s.t. } ij \in g\} \mid$ denote the number of cooperation commitments player i is involved in, and let $l(g) = \frac{1}{2} \sum_{i \in N} l_i(g)$ denote the total number of cooperation commitments in g.

Under the possibility of cooperation commitments, each player's equilibrium effort level is simply a function of the number of cooperation commitments that this player has formed. We denote by

$e_i^*(g)$ the equilibrium effort level of player i under the cooperative network $g \in G$:

$$e_i^*(g) = \begin{cases} e_{BA}\, l_i(g) & \text{if } l_i(g) \geq 1 \\ e^* & \text{otherwise.} \end{cases}$$

The equilibrium payoff of player i, $\pi_i(e_i^*(g), e_{-i}^*(g)) = \pi_i^*(g)$, is given by

$$\pi_i^*(g) = \begin{cases} 2\, e_{BA}\, l(g) + (n - n(g))\, e^* - c(e_{BA}\, l_i(g)) & \text{if } l_i(g) \geq 1 \\ 2\, e_{BA}\, l(g) + (n - n(g))\, e^* - c(e^*) & \text{otherwise.} \end{cases}$$

Our approach to modeling network formation is to dispense with the specifics of a noncooperative game for the formation of a cooperative network and to simply model a notion of what a stable cooperative network is directly (as mentioned below, stable cooperative networks can however be seen as the outcomes of a dynamic network formation process). This static approach was originally taken by Jackson and Wolinsky (1996) and is captured in the following definition. A cooperative network $g \in G$ is *pairwise stable* if $\forall\, ij \in g$, $\pi_i^*(g) \geq \pi_i^*(g \setminus ij)$ and $\pi_j^*(g) \geq \pi_j^*(g \setminus ij)$, and $\forall\, ij \notin g$, if $\pi_i^*(g) < \pi_i^*(g \cup ij)$ then $\pi_j^*(g) > \pi_j^*(g \cup ij)$. The following proposition characterizes the pairwise stable cooperative networks.[7]

Proposition 1 *If n is an even number, the unique pairwise stable cooperative network is such that each player is involved in exactly one cooperation commitment, i.e., it consists of $n/2$ cooperation commitments. If n is an odd number, the unique pairwise stable cooperative network is such that each of $(n-1)$ players is involved in exactly one cooperation commitment and one player is not involved in any cooperation commitment, i.e., it consists of $(n-1)/2$ cooperation commitments.*

The above proposition follows immediately from two results: i) any pair of players such that none of the two players is involved in a cooperation commitment strictly benefits from forming a cooperation commitment, and ii) a player who is already involved in one cooperation commitment is strictly worse off by forming an additional cooperation commitment (see the Appendix for the details of the proof). Consequently, a pairwise stable cooperative network is the outcome of a cooperation commitment-based formation process where players have the discretion to form and sever cooperation commitments one at a time and the formation of a cooperation commitment requires the consent of both involved parties while severance is possible unilaterally. Whether players form or sever cooperation commitments based on the improvement the resulting cooperative network offers relative to the current cooperative network or whether players forecast how their decision to form or sever a cooperation commitment might influence the future evolution of the cooperative network does not play a role.[8]

Without the possibility of cooperation commitments, dominant effort levels do not maximize

[7] In the considered environment, all players are identical ex ante, i.e., before forming cooperation commitments. Consequently, two cooperative networks which share the same architecture but in which the specific players are permuted (these two networks are said to be isomorphic) will lead to the same distribution of equilibrium payoffs. If, when comparing two cooperative networks, all that has changed are the labels of the players, then individual equilibrium payoffs only change to account for the relabelling. Henceforth, the phrase "unique cooperative network" means unique up to a renaming of the agents.

[8] Said differently, whatever their degree of farsightedness, players would form a pairwise stable cooperative network when interacting through the mentioned cooperation commitment-based formation process.

the overall material payoff. We now analyze the extent to which individual incentives to form cooperative networks align with social efficiency, where, again, societal welfare refers to the overall material payoff. Accordingly, a cooperative network $g \in G$ is efficient if $\sum_{i\in N} \pi_i^*(g) \geq \sum_{i\in N} \pi_i^*(g')$ for all $g' \in G$.[9] The following proposition documents a tension between pairwise stable and efficient cooperative networks as soon as the population is composed of strictly more than two players.

Proposition 2 *For a given number of players, an efficient cooperative network comprises at least as many cooperation commitments as the unique pairwise stable network. Therefore, if $n = 2$, the unique efficient cooperative network is identical to the unique pairwise stable cooperative network, i.e., both players form a cooperation commitment. For a given effort cost function, there always exists a large enough number of players such that the efficient cooperative network comprises more cooperation commitments than the pairwise stable cooperative network. For a given number of players $n > 2$, there always exist effort cost functions such that the efficient cooperative network comprises more cooperation commitments than the pairwise stable cooperative network.*

In order to characterize the efficient cooperative networks, we assume that the effort cost function is quadratic, i.e., $c(e) = \mu\, e^2$ with $\mu > 0$. In this case, the dominant effort level is given by $e^* = 1/(2\mu)$ which leads to an equilibrium individual payoff of $(2\,n - 1)/(4\mu)$. The fully efficient effort level is given by $e^{**} = ne^*$ which leads to a fully efficient individual payoff of $n^2 e^*/2$. The individual effort level per cooperation commitment is given by $e_{BA} = 2e^*$ which entails that the equilibrium effort level of player $i \in N$ under the cooperative network $g \in G$ is given by $e_i^*(g) = 2e^*\, l_i(g)$ if $l_i(g) \geq 1$, $e_i^*(g) = e^*$ if $l_i(g) = 0$. Finally,

$$\pi_i^*(g) = \frac{1}{2\mu} \begin{cases} 4\, l(g) + (n - n(g)) - 2\,(l_i(g))^2 & \text{if } l_i(g) \geq 1 \\ 4\, l(g) + (n - n(g)) - \frac{1}{2} & \text{if } l_i(g) = 0. \end{cases}$$

The following proposition characterizes the efficient cooperative networks when $c(e) = \mu\, e^2$.

Proposition 3 *If n is an even number, the unique efficient cooperative network is such that each player is involved in $n/2$ cooperation commitments, i.e., it consists of $n^2/4$ cooperation commitments and each player's equilibrium effort level equals the fully efficient one. If n is an odd number, any network whose total number of cooperation commitments is an integer in the interval $\left[\frac{n(n-1)}{4}, \frac{n^2}{4}\right]$ and $\frac{n+1}{2} \geq l_i(g) \geq \frac{n-1}{2}\ \forall i \in N$ is an efficient cooperative network. If n is an odd number, an efficient cooperative network falls short of full efficiency.*

Interestingly enough, an efficient cooperative network always achieves (almost) full efficiency meaning that the sum of equilibrium efforts always equals (almost) $n^2 e^*$ whereas the sum of equilibrium efforts exerted in a pairwise stable cooperative network equals at most $2\,n\,e^*$.

3 First experiment

The objective of the first experiment is to establish a suitable standard of comparison, to replicate relevant previous findings in order to evaluate the transferability of suggested explanatory

[9] This notion of efficiency is called strong efficiency in Jackson and Wolinsky (1996).

approaches to our environment, and to settle some methodological issues, all necessary to later on *evaluate and explain* the impact of the pairwise commitment mechanism.

Concretely, we implement a repeated version of the effort game described in the previous section with an unknown end where subjects have *no* possibility to form pairwise commitments.[10] We manipulate the availability of information about others' behavior in two different experimental conditions. Our aim here is to test the role of 1) voluntary but costly individualized monitoring, and of 2) repetition with an unknown end as independent mechanisms for solving the public good provision problem. Both of these features have been introduced to more closely mirror the natural decision environment. In voluntary contribution public good provision experiments, subjects usually receive instantaneous costless feedback about the group contribution or/and about all individual contributions.[11] But behavioral information is not automatically and costlessly available in the field. Instead, social actors need to decide whether and how many of their interaction partners they want to monitor, given that monitoring is more or less costly. To the best of our knowledge there is no previous experimental evidence that is evaluating the effects of voluntary costly monitoring on the solution of social dilemma problems. A second highly artificial feature of standard voluntary contribution public good provision experiments is the implementation of finitely repeated games with a commonly known end-point. The rationale underlying this operationalization is to prevent repeated game effects and the theoretically resulting plethora of possible equilibria. Normann and Wallace (2005) have only recently provided some first comprehensive evidence concerning the actual impact of different termination rules in a repeated Prisoner's Dilemma experiment. They are the first to compare treatments with a known finite end, with an unknown end and with two versions of a random termination rule in a *unified* framework. They find that the different termination rules do not significantly affect cooperation rates, which suggests that repeated game considerations are behaviorally irrelevant. Our objective in the first experiment is to test whether Normann and Wallace's result can be replicated in our environment, showing that repetition with an indeterminate end point, while closely resembling the realistic absence of common knowledge about the end of an interactive relationship, does not induce supergame play. Answering this question is a necessary precondition to determine the best-suited implementation to evaluate the pairwise commitment mechanism's relative capacity to *sustain* cooperation, as the latter is obviously more difficult to discern *objectively*, once artificial end-game effects - which are a direct implication of the publicly known finite end implementation - need to be accounted for.

Finally, the first experiment is aimed at establishing whether and to what extent subjects who are also concerned about the payoffs of others do play a role in our environment, which serves to determine appropriate explanatory approaches to later on interpret the results of the second

[10]The effort game corresponds to a non-linear pure public good game. Henceforth, both terminologies will be used. We decided to rely on a non-linear setting in order to later on more cleanly disentangle the pairwise commitment mechanism's impact while at the same time giving it best chances. This is based on previous experimental evidence which suggests that contributions in non-linear public good experiments tend to be closer to the predicted level than in linear settings.

[11]Some studies have investigated the impact of instantaneous costless individualized contribution feedback: Weimann (1994) and Croson (1998) find no difference between aggregate and individual contribution feedback, while Sell and Wilson (1991) find that individualized feedback increases contributions in late periods. Cason and Khan (1999) have investigated 'imperfect monitoring', in the sense that subjects only after every sixth period receive - again automatic and costless - information about aggregate group contributions and find no difference in the observed contribution levels contingent on whether monitoring is perfect or not.

experiment.

3.1 Experimental design

In the course of a session each experimental subject goes through a *fixed but unknown* number of repetitions of the effort game, in which he interacts with the same three other subjects (*partner design*). In each repetition, a subject's effort has to be an integer in $\{0, 1, \ldots, 12\}$, where the individual cost function is given by $\mu\, e^2$ with $\mu = 1/4$. This implementation corresponds to the Voluntary Contribution Mechanism (VCM) which has been introduced by Isaac, Walker, and Thomas (1984).

We consider two experimental conditions. In the first experimental condition, called the *no feedback condition* (in the following abbreviated as $NoFeedback$), subjects do not have any possibility to observe or infer their partners' behavior. At the conclusion of a repetition, subjects are neither informed about their resulting payoff (which would allow them to reconstruct their partners' aggregate effort level) nor can they directly monitor the individual effort exerted by any other person in their group or in any other group.

In the second experimental condition, called the *baseline condition*, each subject has access to information about his partners' behavior, as 1) subjects are informed about their payoff at the conclusion of each repetition, and 2) each subject has the possibility to directly monitor the *individual* effort levels of some or all of his partners if he wishes to do so. We assume that monitoring is costly and we consider two levels of monitoring cost: a low level and a high level.

Table 1 summarizes the experimental design of our first experiment, while we provide the implementation details of the two experimental conditions in the following.

Treatment condition	Possibility to form commitments	Own payoff feedback	Monitoring of others' individual behavior	Monitoring cost
$NoFeedback$	no	no	no	—
$BaseLow$	no	yes	yes	low
$BaseHigh$	no	yes	yes	high

Table 1: Experimental design of the first experiment.

3.1.1 No feedback condition

The actual number of repetitions (periods) or the duration of a session was *not known* to any of the subjects. Instead, subjects were invited for three hours and informed that they will interact for between 55 and 80 periods with the same three other subjects. The actual number of periods in this condition was 65 which translates into an effective session duration of slightly more than one hour.

In a given period, each group is composed of a subject labelled 'person A', a subject labelled 'person B', a subject labelled 'person C', and a subject labelled 'person D'. Labels were randomly assigned to subjects at the beginning of the session and kept for its entire duration. In each period all subjects simultaneously chose an effort level between 0 and 12 which they entered and confirmed on a 'decision screen'. Subjects were instructed that their total payoff in each period is given by the sum of the four effort levels minus the cost of their own effort. All possible payoffs resulting from all possible combinations (of own effort, sum of the three other efforts) were provided in a payoff table, which permitted a comprehensive overview over the entire range of possible payoffs. This guaranteed that subjects had complete information about the relationship between decisions and payoffs applying to themselves and all others. At the end of each period, a 'result screen' was displayed to every subject on which he was reminded of the effort level he had chosen but did not obtain any additional information. After the 65th period, each subject was informed about the payoffs he had accumulated over the course of the session.

Before dismissing subjects we asked them to answer an incentivized post-experimental questionnaire in which we tested their ability to point out the dominant strategy of the one-shot version of the effort game assuming a self-interested subject.[12] From now on, we simply refer to the dominant strategy under the assumption of selfishness as to the equilibrium effort level (contribution).

3.1.2 Baseline condition

The implementation of the baseline condition has been identical to that of the no feedback condition, except for the following aspects: 1) the actual number of periods was 60 in both the low and the high monitoring cost treatments, which again translates into an effective session duration of slightly more than one hour, and 2) at the conclusion of a period, each subject was - on his 'result screen' - not only reminded of his own effort level but was now also informed about his resulting payoff. Additionally, each subject had the possibility to monitor the individual effort levels of each of his partners by clicking on three accordingly labelled boxes (which when clicked on revealed the requested information). A fixed cost of 0.3 (respectively 2) experimental points per displayed effort level in the low (respectively high) monitoring cost treatment was automatically subtracted from the subject's displayed period payoff.

The parametrization has been chosen such that monitoring costs were effectively negligible in the low monitoring cost treatment, but considerable in the high monitoring cost treatment. If subjects in a given period of the low monitoring cost treatment buy the maximal reasonable amount of information, i.e., if they monitor two of their interaction partners which allows them to deduce their third partner's effort, their monitoring cost amounts to 8% of the equilibrium payoff. Instead, subjects need to invest 57% of the equilibrium payoff to buy the same amount of information in the high monitoring cost treatment.

[12]Subjects received 1 euro for correctly providing the dominant strategy. They were informed about the remuneration on their questionnaire sheets. The questionnaire consisted of two parts: In the first part, we asked subjects whether or not they think that there is one effort level which maximizes their own payoff for all efforts that the other three persons can choose and if so, to name it. Only those subjects that answered the first part correctly received a second part in which they were asked to indicate whether they have chosen the dominant strategy in all 65 periods and if not, why not.

3.1.3 Practical procedures

We ran one session with 6 groups of four subjects in the no feedback condition and one session with 6 groups of four subjects in each of the two monitoring cost treatments of the baseline condition. Consequently, we have collected six independent observations based on 24 subjects in NoFeedback and six independent observations based on 24 subjects in both BaseLow and BaseHigh.

The experiment has been conducted using the computerized network of the Max Planck Research Laboratory at Jena, Germany. In total 30 subjects were invited for each session, 24 of which (six groups of four subjects) took part. All 72 subjects participating in the first experiment were students of Jena University. Additional subjects were invited to ensure that only subjects that fully understood the procedures (tested by a pre-experimental questionnaire) participated. Participating subjects were reading for humanities as well as medical, social, mathematical/physical and environmental/life science degrees. Some subjects had participated in economic experiments before, but none had taken part in an experimental setting similar to the one reported here.[13] The gender composition in the first experiment was approximately balanced with 53% male subjects.

Subjects interacted via computer terminals, which were visually isolated from each other. Communication other than through taken decisions was not allowed. Subjects were fully informed about the rules of the game by written instructions which were rendered public knowledge by reading them out loudly. The instructions were *neutrally formulated* meaning that subjects chose numbers rather than effort levels.[14] In addition to the written instructions in which benefits and costs were provided in functional form, the parametrization was chosen in a way that - as mentioned - allowed us to equip each subject with a payoff table.[15] Moreover, each subject was provided with a 'history sheet' for voluntary use, making it easy to keep track of own (NoFeedback) and own/others' decisions and outcomes (in both Base treatments) during the experiment if a subject wished to do so. In NoFeedback subjects were notified in the instructions that they could ask the experimenters after the end of the session to provide detailed information about the outcomes of each period. This was done to make sure that subjects trusted that they indeed interacted with three other subjects even though they received no feedback during the session.

In order to disable beliefs that the experimenters would randomly stop the experiment we informed subjects in the instructions that the actual number of periods is fixed. We rendered this credible by sticking a poster on the wall - before the start of the session - in sight of all subjects, which had the actual number of periods written on its back. We turned the poster at the end of the session to prove that the number of periods had been fixed ex-ante.

In their invitation to the experiment subjects were informed that participation was conditional on questionnaire performance and that there is no show-up fee.[16] Subjects, whose questionnaire results indicated that they had not sufficiently understood the game (those that made two or more mistakes), were replaced and paid the minimal compensation of 2 euros for answering the questionnaire. The questionnaire comprised four questions checking subjects' understanding of

[13] We have excluded highly experienced subjects which had taken part to 12 or more experiments.

[14] Also there was no reference to 'groups' in the instructions or on the screens. Instead we spoke of interacting persons. See **??**.

[15] Payoff tables were printed in poster size (A3) and cellar-taped on the partition screens of each cubicle so that they were comfortably readable by subjects. See **??**.

[16] Subjects were invited using ORSEE (Greiner, 2003).

the payoff calculation and respectively their ability to read the payoff table. The questionnaire was identical in the two experimental conditions. Overall, 74% of all invited subjects made no mistake in the questionnaire. 78% of those subjects that finally participated had made no mistake in the questionnaire, the rest made one mistake. Participating subjects received 4 euros if they had answered the questionnaire correctly and 2 euros if they made one mistake. The amount earned for the questionnaire served as an initial endowment to cover potential later losses. As these could not generally be excluded (even though they were very unlikely), all subjects were necessarily required to sign a letter of agreement before the start of the experiment informing them about the possibility of losses and asking them to cover losses from the money earned for the questionnaire.[17]

Subjects' earnings were tallied in 'laboratory points' and converted to euros at the end of the experiment. Conversion rates from points to actual euro earnings were set at 0.015 euros per point, again irrespective of the experimental condition and the monitoring cost treatment.

3.2 Predictions

Given our parametrization, the equilibrium effort level, in the one-shot version of the effort game, equals 2 units which leads to an equilibrium payoff of 7 experimental points. The efficient effort level equals 8 units which translates into a payoff of 16 experimental points for each individual.

No information is spread among subjects in the no feedback condition. Therefore, all approaches assuming self-interested players predict that actors will revert to the equilibrium effort of the stage game. In the baseline condition, regardless of the level of monitoring costs, the maximal number of repetitions of the stage game is public knowledge. Hence, the unique subgame perfect equilibrium of the repeated effort game with unknown end is for all players to free-ride in all periods. One possibility to rationalize some measure of cooperation in the repeated version of the effort game with feedback is to alter the strict preference relations on outcomes (Radner (1980) [Collusive Behavior in Non-Cooperative Epsilon-Equilibria in Oligopolies with Long but Finite Lives, JET, 22, 136-154]). Another possibility is to perturb the strategy sets of the players such that at least one of the players is restricted to mixed strategies that with a strictly positive probability chooses a pre-specified strategy (Kreps, Milgrom, Roberts, and Wilson (1982) [Rational Cooperation in the Finitely Repeated Prisoners' Dilemma, JET, 27, 245-252]). By comparing the observed level of cooperation in the no feedback and baseline condition we investigate the behavioral relevancy of these approaches.

3.3 Results

In all following analyzes we restrict attention to periods 1 to 50 for which we are sure that end game considerations have been irrelevant as subjects knew that they would go through at least 55 periods.

Figure 1 depicts the average effort paths over the first 50 periods for the NoFeedback condition as well as for both baseline treatments. We only display the relevant part of the action space between the equilibrium effort of 2 units and the efficient effort of 8 units. Table 2 summarizes the

[17] By signing they also agreed to cover losses exceeding this amount by incomes from future experiments (which in fact never applied). See footnote??. Two subjects did not agree to sign and consequently could not take part to the experiment.

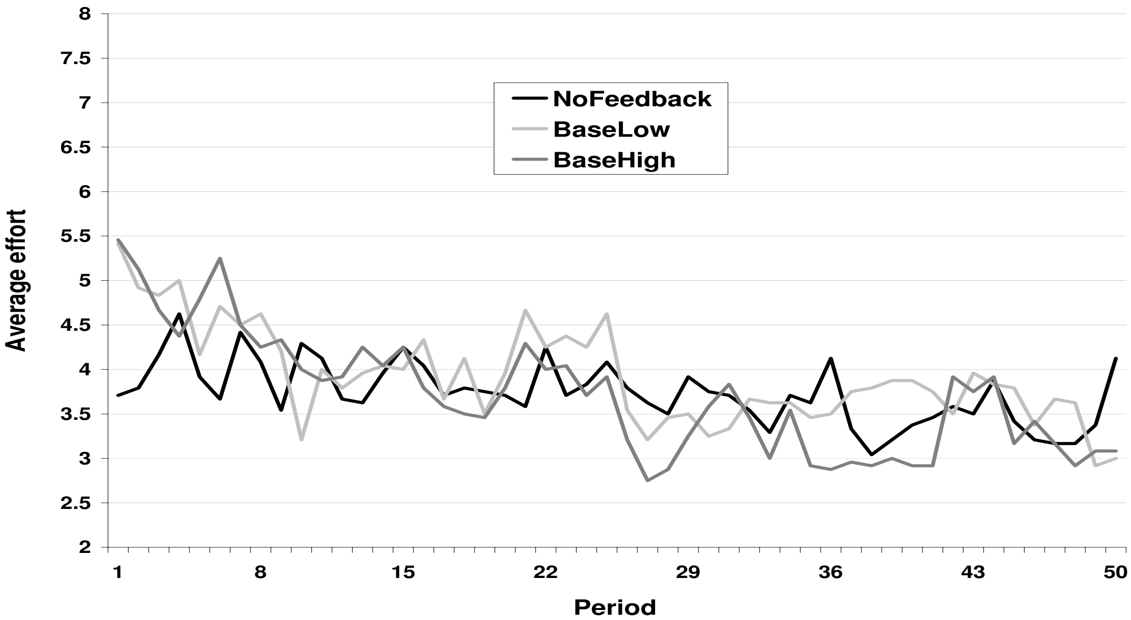

Figure 1: Average effort paths in the first experiment.

findings at the aggregate level by providing, for each condition/treatment, the average effort level and the associated *cooperation rate*. The latter states the average effort level as a fraction of the efficient level (normalized to the range between 2 and 8 units).

	Average effort (effort's standard deviation): First 10 periods	Last 10 periods	Cooperation rate All periods
NoFeedback	4.02 (0.36): 33.68 %	3.48 (0.30): 24.62 %	3.73 (0.35): 28.90 %
BaseLow	4.56 (0.60): 42.64 %	3.57 (0.35): 26.20 %	3.92 (0.54): 32.00 %
BaseHigh	4.68 (0.48): 44.58 %	3.30 (0.39): 21.59 %	3.71 (0.67): 28.53 %

Table 2: Average effort levels and cooperation rates in the first experiment.

Averaging across all periods and groups, subjects exhibit a cooperation rate of 28.9% in NoFeedback. This replicates previous findings according to which subjects cooperate beyond the equilibrium level even *without any behavioral feedback* (Sell and Wilson, 1991; Wilson and Sell, 1997). When behavioral feedback becomes available in our baseline condition the results look very similar, irrespective of the monitoring cost, with the exception that initial cooperation rates are much higher. But this difference vanishes almost immediately, which gets reflected in a strong initial decrease in cooperation rates in both baseline data sets which wears out over time. In consequence, the aggregate average cooperation rates overall differ only very mildly between the NoFeedback and the two treatments under the baseline condition. This is a first clear indication against the conjecture that repetition with an indeterminate length - combined with behavioral feedback - suffices to induce and sustain cooperative equilibria by enabling subjects to strategically exploit repeated game effects.

We resort to a linear mixed-effects model[18] to estimate the average effort level, and its evolution in the different treatments, in order to pin down the significance of the mentioned first impressions.[19] The fixed effects consist of an intercept, the continuous variables $Period$ and $1/Period$ (which is included to capture the decreasing marginal change in the effort levels suggested by the data) and the interaction effect $NoFeedback * 1/Period$. The random effects only appear at the individual level, and consist of random intercepts. The assumption of a within-subject first-order autoregressive form of the error term provides a substantially better fit of the data than the independent errors model (the estimated autocorrelation is 0.236).[20] We additionally model non-constant variance by considering an exponential error model for the interindividual variability. The results of the regression are displayed in table 3.[21]

	Value	p-value
Intercept	4.32	< 0.01
1/Period	1.20	< 0.01
Period	-0.02	< 0.01
NoFeedback*1/Period	-1.77	< 0.01
Random Effects: Intercept = 1.38		
Number of observations: 3600		
Number of individuals: 72		

Table 3: Results of the statistical analysis of efforts in the first experiment.

First, the regression results confirm that there is no significant difference between the two baseline treatments, as the cost of information has no significant effect on the estimated average effort level (which is why we will in the following pool the data and jointly refer to the two treatments as *Base*). Secondly, the estimated initial effort level in the first period of the NoFeedback condition is with 3.73 units significantly lower than the effort level estimated for the Base condition (5.49 units). The statistical analysis suggests a significant decrease in the effort level in Base, where the level of cooperation drops to 3.39 units in the last period (it drops to 3.36 units in the last period of NoFeedback). The significant impact of the variable $1/Period$ underlines that the effort level deteriorates at a decreasing marginal rate in Base. Starting with period 20, there remains no difference in the average cooperation level or its trend between NoFeedback and Base.

[18] In each session, subjects interact for several periods. We have therefore collected repeated measurement data, i.e., data where subjects have multiple measurements over time. Analyzing these data requires recognizing and estimating variability both between and within subjects. Mixed-effects models serve this purpose.

[19] Acceptance or rejection of the null hypothesis is always based on a 1 percent level of significance.

[20] Since our observations are taken longitudinally on the same subjects, within-subjects errors are likely to be auto-correlated.

[21] A normal plot of the within-subjects standardized residuals indicates that the assumption of normality for the within-subjects errors is plausible. The considered model is the outcome of a selection process which started with a full model including all possible interaction effects (except for Period*1/Period) and random intercepts also at the group level, and then sequentially dropped insignificant variables (such as the fixed effect *monitoring cost*) on the basis of likelihood ratio tests. Even though a 2-sided censored Tobit regression is more appropriate as our dependent variable belongs to the set $\{0, \ldots, 12\}$, we conjecture that it would not alter our results qualitatively and that even quantitative differences would be minor. Indeed, only 6.41% / 8.58% / 7.17% of our data are censored in the NoFeedback/BaseLow/BaseHigh treatments respectively.

One possible conjecture to explain that the actual cost of monitoring remains without impact on the observed average cooperation rates is that subjects did not actually use the opportunity to monitor their partners but instead solely relied on aggregate/average information which they could reconstruct from their payoff feedback for free. But this conjecture is not confirmed by our data for the low monitoring cost treatment. In fact, each subject on average monitored 21.33 effort choices of his partners over the course of the 50 periods when the cost of monitoring was low. This amounts to 21.33% of the maximal reasonable monitoring activity (where a given subject monitors the effort exerted by two of his partners and deduces the third partner's behavior in each period). Consequently, we observe an average monitoring activity which is far from negligible. Moreover, the average is hiding a large heterogeneity between subjects. Even though, 42% of the subjects never monitored any of their partners' effort choices, and 12% looked up less than 10 effort choices, another 12% of the subjects monitored 90 effort choices, which amounts to almost the maximal reasonable amount. Opposed to that - and as intended - subjects indeed refrained from monitoring in the high monitoring cost treatment. The average total number of requested information pieces concerning others' behavior amounts to merely 3.83. Concluding, the observed missing impact of the monitoring device on the aggregate cooperation rates cannot be attributed to its general rejection. In fact, some subjects make extensive use of it, at least when its costs are negligible. But obviously, individualized monitoring does not seem to help subjects to sustain high cooperation rates in the *aggregate*.

Up to now we argued at the highest level of aggregation. Though suitable to provide some first impression the aggregates do not adequately capture the structure of our data, given the high estimated random intercept. But before moving to individual behavior, we will first take the intermediate step and look at the level and evolution of cooperation in different groups. In the NoFeedback condition we firstly observe that groups are relatively homogeneous with respect to their average level of cooperation (see, table 13 in the Appendix for details). The lowest group cooperates at an average rate of 11.3%, while all other groups cooperate at average rates between 30 and 40%. Secondly, in half of the groups cooperation rates are stable over time (Groups 1, 3 and 4). Surprisingly, one group - irrespective of the lack of feedback - increases its average effort provision over time (Group2). In the remaining groups cooperation rates exhibit a decreasing trend which - given the stability of the majority of the other groups - gives the grounds for the observed slight decrease in the aggregate trend.

In comparison, groups are relatively more heterogeneous with respect to the average level of cooperation in the Base condition (see, tables 14 and 15 in the Appendix for details). Across different groups average cooperation levels are spread from 13.9% to 65.4% in BaseLow (and respectively from 7.2% to 46.5% in BaseHigh). Different from groups under the NoFeedback condition, cooperation breaks down over time in the majority of groups in the two baseline treatments, viz., in 2/3 of all groups in BaseLow (Groups 1, 3, 4 and 6) and in all groups in BaseHigh (the remaining groups in BaseLow cooperate at a stable rate).

Concluding, group behavior is more homogeneous and cooperation rates more stable when behavioral feedback is not available, while the accessability of aggregate and individualized behavioral information induces a larger dispersion in group performance and a decreasing trend in cooperation rates for the majority of the groups.

We now take a look at individual behavior. We first do so in the NoFeedback condition, which provides an environment which is best-suited to identify different player types (if appropriate) as subjects do not have any incentive to strategically conceal their type. Indeed, we find that different types play a role. Only 46% of the subjects are purely self-interested, in the sense that they exert an average effort level at or just slightly above the equilibrium (see, table 4), while 21% of the subjects provide a very high average effort over the 50 periods and can be referred to as cooperators.[22] The third main category captures 25% of the subjects and groups those that exhibit a largely unsystematic effort provision pattern (signified by a very high standard deviation).[23] By squaring this classification with the results from the post-experimental questionnaire, we obtain that all self-interested subjects were able to state the dominant strategy of the one-shot game. Interestingly, 80% of the cooperators were also able to do so. This suggests that their cooperative behavior does *not* stem from confusion but from a preference for cooperation/efficiency, which they also expressed in the questionnaire. Opposed to that only half of the 'unsystematic types' were able to identify the dominant strategy. This indicates that their unsystematic behavior is a direct expression of their inability to grasp the incentive structure underlying the decision situation.[24] Concluding, we find that the aggregate cooperation rates are not driven by the unsystematic types, as their exclusion does not modify the aggregate outcomes. Instead, the observed behavior is best explained by the interaction of self-interested subjects and cooperators.

	Relative frequency (frequency)	Av. effort (effort's std.)/Coop. rate	Able to state dominant strategy
Non-cooperators	45.8% (11)	2.27 (0.69): 4.4%	100%
Cooperators	20.8% (5)	6.50 (1.29): 74.9%	80%
Unsystematic	25.0% (6)	3.98 (2.54): 33.1%	50%

Table 4: Subjects' types in NoFeedback.

The central qualitative finding in the baseline condition is that the fraction of subjects that exhibit a stable cooperation rate is halved compared to the NoFeedback condition. The same holds for the fraction of individuals that are exerting efforts at a very low and at a very high level. The move away from the extremes is attributable 1) to self-interested players that are now cooperating strategically to induce others to also cooperate, and 2) to conditional cooperators that - different from the NoFeedback condition - are now able to adapt their initially too optimistic beliefs. This suggests that once information about others' behavior is accessible, strategic considerations do

[22]Those subjects could be either conditional cooperators with optimistic beliefs about the behavior of others or unconditional cooperators.

[23]Two subjects have not been classified. They cooperate at an intermediate and rather stable level. These findings are qualitatively in line with previous findings which suggest that about half of the population in standard social dilemma experiments are conditional cooperators (i.e., subjects that cooperate if others cooperate as well), one third is purely selfish, while the rest displays unusual, not easily identifiable patterns (Fischbacher, Gächter, and Fehr, 2001; González, González-Farìas, and Levati, 2005).

[24]Clearly, not understanding the underlying incentives is different from not understanding the rules of the game.

come into play to some extent, but subjects are clearly unable to exploit repeated game effects in a way that would allow them to sustain cooperation rates at a high level over time.

3.4 Discussion

Our first experiment establishes a 'base rate of cooperation' moderately beyond the equilibrium level of the stage game. By comparing the results of two appropriately designed experimental conditions we show that this base rate cannot be attributed to self-interested players' strategic exploitation of repeated game effects. Although strategic considerations do play some role, they only induce a higher initial cooperation rate, which cannot be maintained but starts to break down immediately. Consequently, our results establish that repetition with an unknown end does not provide a mechanism which is *behaviorally* sufficient to sustain cooperative outcomes. Rather than to supergame play, we show that the observed base rate of cooperation seems to be attributable to the interaction between (un/conditional) cooperators and self-interested players, while confused subjects (though present in the population) do not drive the results.

Taken together these results provide a clean methodological foundation and a straightforward standard of comparison to now evaluate the appeal and the efficacy of the pairwise commitment mechanism.

4 Second experiment

The objective of the second experiment is to evaluate the relative efficacy of the pairwise commitment mechanism to induce, and to sustain cooperation contingent on the cost of individual monitoring. As before, the design has been chosen to reflect essential features of the natural environment as closely as possible.

4.1 Experimental design

In the second experiment we implement one experimental condition, called the *commitment condition.* It directly builds up on the implementation of the baseline condition in the first experiment, the only difference being that we now introduce the possibility to form pairwise cooperation commitments by which both involved subjects mutually restrict their respective action space in subsequent periods of the effort game. In our implementation, commitments are: i) *voluntary* meaning that commitment formation is optional rather than mandatory in order to prevent an artificial demand effect; ii) *mutual* and *consensual* meaning that both involved subjects restrict their respective action spaces if and only if they agree to do so; and iv) *binding* meaning that commitments - once formed - are externally enforced. Table 5 summarizes the design of the second experiment.

Per commitment, both involved subjects raise the lower bound of their respective action space by 4 units, i.e., subject i's action space is given by $\{4\,l_i, \ldots, 12\}$, where $l_i \in \{0, 1, 2, 3\}$ is the number of formed commitments. Consequently, we exogenously induce commitments which are efficient at the pair level. This allows us to retain a greater level of control by giving best chances to the emergence of pairwise stable networks.[25] In the laboratory the bindingness of commitments is enforced by

[25] As indicated in our introduction, the compensation mechanism suggested by Varian (1994) may be one possibility

Treatment condition	Possibility to form commitments	Payoff feedback	Monitoring of others' individual behavior	Monitoring cost
CommitHigh	yes	yes	yes	high
CommitLow	yes	yes	yes	low

Table 5: Experimental design of the second experiment.

means of the software which does not permit effort entries smaller than the minimal effort a subject has committed to given the number of cooperation commitments that he has formed.[26] Different from the baseline condition, subjects actually went through only 55 periods in the course of a session, which translated into an effective session duration of approximately 2 hours.

Session structure

For the first 10 periods, the commitment condition fully parallels the baseline condition meaning that subjects have no possibility to form pairwise commitments. This guarantees that subjects first gain some experience with the baseline effort game which induces the normal amount of defection and potential frustration before the commitment mechanism is introduced.

At the end of the $10th$ period, subjects are asked to state whether they would like to modify the up to then prevailing empty network by forming commitments or not. The commitment negotiation stage (described below) is entered only *on demand*, viz. if at least two subjects state that they want to form cooperation commitments, which is a direct consequence of the assumption that cooperation commitments cannot be formed unilaterally but instead require the concerted initiative of pairs of players.

Independently of whether subjects entered the negotiation stage and finalized a set of pairwise cooperation commitments or whether they decided to keep the empty network (skipping the commitment negotiation stage), they afterwards move on with the effort game, given the potentially formed pairwise commitments, for another block of 10 periods. This introduces a realistic degree of inertia, which reflects the fact that cooperation agreements in the field are effective for a while before they can be resolved. Furthermore, this gives subjects a better chance to understand the impact of a given network.

After each additional block of 10 periods, subjects have an opportunity to *renegotiate* the network. Concretely, subjects are asked to state whether they would like to modify the existing cooperative network at the end of the k^{th} repetition, $k \in \{10, 20, 30, 40, 50\}$.

to endogenize the commitment level. This mechanism has currently only been tested experimentally for the two-player Prisoner's Dilemma (Andreoni and Varian, 1999; Charness, Fréchette, and Qin, 2005). When applied to our setting, the determination of the optimal effort level interacts with the determination of the incentive-compatible network. For that reason it makes sense to proceed step by step in order to be able to disentangle effects. In this sense, our implementation provides a natural starting point.

[26]Each subject is reminded of the current network and the resulting action space restriction applying to himself and the three persons he is interacting with on his 'decision screen'.

Optional commitment (re)negotiation stage

If subjects enter the optional commitment negotiation stage at the end of the first block of 10 periods, they do not have any commitments yet, i.e., the initial network is the empty one. Starting from here the four subjects can sequentially form pairwise cooperation commitments according to an exogenously fixed but ex ante unrevealed *order of play*.[27] Subjects were instructed that the order of play is such that each possible pair is called exactly once. As there are six possible pairs, the order of play is given by the six pairs called in six consecutive *steps*. We denote a complete move through the order of play as a *sequence*. At the beginning of each step all subjects are informed about which pair is called on to decide. Within the called pair, decisions are stated sequentially, where the subject deciding second is informed about the first subject's decision before stating his own. A cooperation commitment between a called pair is formed if and only if *both* subjects agree on its formation. At the end of a step, all subjects are informed about the then prevailing network, where subjects that have not been involved in the currently deciding pair are just informed about the outcome of the pairwise decision.[28] Afterwards, a new step starts and the next pair determined by the order of play takes a decision.

At the end of a sequence, all four subjects are asked to state whether they want to modify the then prevailing cooperative network or not. If all four state that they do not want to make any modifications, the first stage is terminated, meaning that the network is definitive and the subjects move on with the next block of ten periods of the effort game. This possibility of ending the commitment negotiation stage corresponds to *termination rule 1*. Otherwise, an additional sequence is initiated in which all possible pairs are recalled according to a different order of play than the one applied in the first sequence. Subjects then have the possibility to reconsider their respective relations. Concretely, it is possible to form commitments that have not been formed yet, to delete commitments previously formed, or to keep all relations unchanged.

After all possible pairs have been recalled, the following additional termination rule applies: if a network has not actually been modified in the course of a recall-sequence, i.e., the network at the end of the sequence is the same as at the end of the previous sequence, the negotiation stage is terminated, the prevailing network is definitive and subjects move on with the next block of ten periods of the effort game. This second possibility of ending the negotiation stage corresponds to *termination rule 2*. If the network has been modified, every subject is again asked to state whether he wants to change the then prevailing network or not. In this case, the rules specified above apply. Overall, the maximal number of sequences is five. Consequence of which is that, after the fifth recall of all possible pairs, the negotiation stage is terminated and subjects move on with the effort game. This last possibility of ending the negotiation stage corresponds to *termination rule 3*.

If subjects enter the commitment negotiation stage in later phases of the game to *renegotiate* a previously formed cooperative network, the applied protocol is identical, except that the initial network may or may not be given by the empty network.[29]

[27] The implemented orders of play are provided in table **??** in the Appendix. They have not been revealed to subjects ex ante in order to induce myopic behavior which in turn gives best chances to the formation of pairwise stable networks.

[28] Of course, individual decisions can be straightforwardly inferred if a cooperation commitment is formed.

[29] And, if in $k > 10$ the prevailing network is not given by the empty network any longer, the commitment negotiation stage is entered also if only one player asks for modification, as long as this player is connected in the

In the course of the commitment (re)negotiation process subjects receive the following feedback on their screens: 1) at any time of the (re)negotiation process, all persons have a display of the currently prevailing cooperative network on their screens, 2) all persons are informed at the beginning of each step, which pair is called to decide, 3) within a called pair, both subjects are informed about their respective decisions, while 4) subjects which are not part of the currently deciding pair are only updated about the outcome (and not about individual decisions) at the end of the step.

4.1.1 Practical procedures

Overall, we ran three sessions in each of the two commitment treatments. In both treatments, the first session has been run with 5 groups, the second session with 3 groups and the third session with 4 groups. Consequently, we have collected 12 independent observations based on 48 (12×4) subjects in each of the two treatments. The gender composition has been balanced with 53% male subjects.

The practical procedures were identical to those of the first experiment except for the following modifications: The pre-experimental questionnaire has been extended to ten questions in which we - next to the understanding of the payoff calculation - checked subjects' understanding of the commitment negotiation procedure. The questionnaire was identical in the two commitment treatments. Overall, 67% of all invited subjects and 74% of those subjects that finally participated made no mistake in the questionnaire, while 23% made one mistake. Due to no show-ups we could not replace three subjects with two mistakes.

Each subject was provided with four copies of the baseline payoff table representing the four possible cases (of zero, one, two and three own cooperation commitments). In the four copies those parts of the action space that a subject effectively excluded by forming the respective number of commitments were blackened. The idea here was to visualize the action space restriction implied by commitment formation.

4.2 Predictions

In the four player setting that we are implementing in the laboratory, there are 64 different possible networks. These can be assigned to eleven different architectures. Table 6 provides one example network for each architecture, where all networks belonging to the architecture are obtained by going through all possible permutations of player labels.

Given our parametrization (taken over from the baseline condition), the unique set of pairwise stable networks comprises the three possible 2-link networks, denoted by g^2. Here each of the four players forms exactly one commitment, exerting an effort level of 4 units which leads to an equilibrium payoff of 12 points. Instead, the fully efficient network architecture is given by the circle, g°, where each player forms exactly two commitments, exerting an effort level of 8 units, which leads to a payoff of 16 points.

existing network. This accounts for the fact, that commitments can be deleted unilaterally.

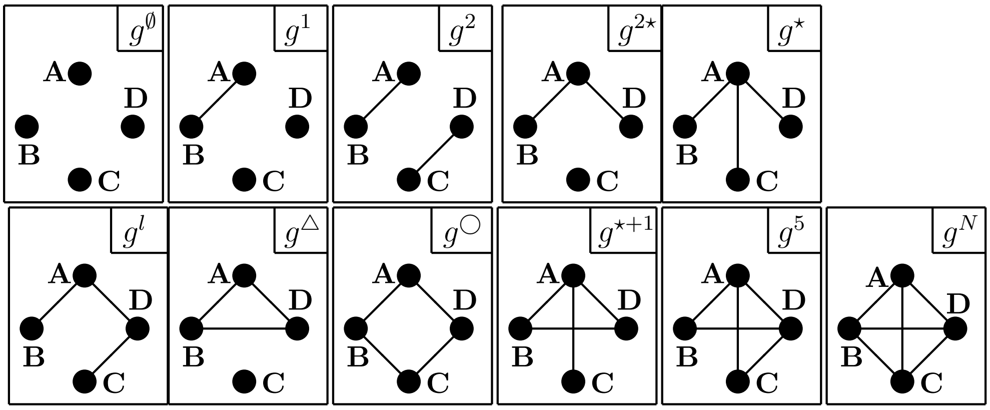

Table 6: Example network structures for the 11 possible architectures.

4.3 Results

Figure 2 depicts the average effort paths for the two commitment treatments, and puts them into perspective to the baseline effort path (which pools the data from BaseLow and BaseHigh). Table 7 summarizes the average aggregate results for the two commitment treatments.

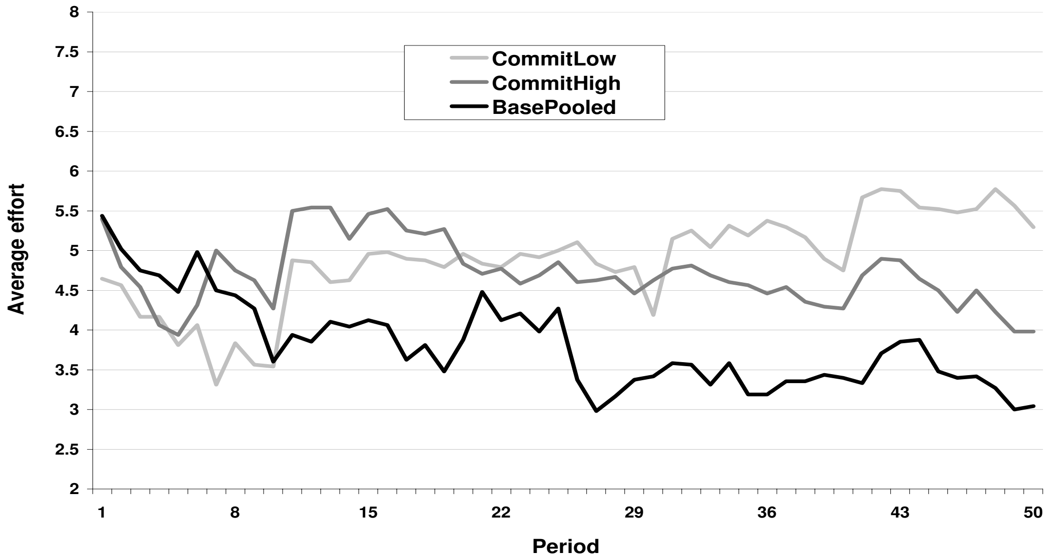

Figure 2: Average effort paths in the second experiment.

	Average effort (effort's standard deviation): First 10 periods	Last 10 periods	Cooperation rate All periods
CommitLow	3.97 (0.44): 32.78 %	5.51 (0.29): 58.52 %	4.87 (0.59): 47.84 %
CommitHigh	4.57 (0.44): 42.81 %	4.44 (0.32): 40.59 %	4.71 (0.42): 45.14 %

Table 7: Average effort levels and cooperation rates in the second experiment.

The pairwise commitment mechanism is effective. The observed aggregate average effort levels clearly exceed those observed under the voluntary contribution mechanism in the baseline condition. In fact, the aggregate effort levels lie - on average over all 50 periods - even beyond the equilibrium level of 4 units (translating into a cooperation rate of 33.33%) which is induced by the formation of pairwise stable networks. We observe a large immediate increase in the aggregate average effort already in the very first period after the pairwise commitment mechanism has been introduced (Period 11). Although the overall aggregate average effort levels are similar between the two commitment treatments, we observe a clear difference in the evolution of cooperation contingent on the cost of monitoring. In the low monitoring cost treatment (CommitLow), the commitment mechanism not only avoids a decrease in the exertion of effort over time but moreover leads to a continuous increase from an initial level in Period 11 which is comparatively lower than that observed in the high monitoring cost treatment. In contrast, the initially high average effort level cannot be sustained in the high monitoring cost treatment (CommitHigh). Here aggregate cooperation rates deteriorate over time.

We recur to a series of regressions to discern the significance of the observed differences and to more clearly characterize the dynamics. Concretely, we jointly estimate the average effort under the voluntary contribution mechanism (based on the data of the baseline treatments of the first experiment) and the average effort under the pairwise commitment mechanism.

In a first step, we restrict attention to the first 10 periods. These have been identically implemented across the two experiments, viz., by relying exclusively on the voluntary contribution mechanism without permitting the formation of cooperation commitments. The originally considered complete linear mixed-effects model to estimate the average effort level could be reduced to a model where the factor $Baseline$ (which distinguishes between the two experiments) and any interaction effect involving it is not significant. Accordingly, there is no significant difference in average effort provision in the first 10 periods of the baseline as compared to the commitment treatments. Hence, we can rule out general differences between the two settings.[30] Based on these results, we can in the second step now safely restrict our focus to the last 40 periods, in which the commitment mechanism has been available in the commitment but not in the baseline treatments. Again we make use of a linear mixed-effects model to estimate the respective average effort levels. The fixed effects consist of an intercept, the continuous variable $1/Period$, the variables $Baseline$, $Infocost$ and the interaction effects $Baseline * 1/Period$, $Infocost = high * 1/Period$ and $Baseline * Infocost = high * Period$. The random effects appear at the individual and group level, and consist of random intercepts.[31] The results of the regression are displayed in table 8.

According to the results of the regression, the estimated average effort level in the $11th$ period of CommitLow is with 4.51 units very similar to effort levels that are estimated for the baseline

[30]The latter may have resulted from the fact that subjects in the commitment treatments were informed about the introduction of the commitment mechanism at the end of the $10th$ period.

[31]Assuming a within-subject first-order autoregressive form of the error term again provides a substantially better fit of the data than the independent errors model (the estimated autocorrelation is 0.458). We additionally model non-constant variance by considering an exponential error model for the inter-individual variability. A normal plot of the within-subjects standardized residuals indicates that the assumption of normality for the within-subjects errors is plausible. The considered model is the outcome of a selection process which started with a full model including all possible interaction effects. Again, almost none of our data are censored: these are 3.31% / 2.62% in the CommitLow/CommitHigh treatments respectively.

	Value	p-value
Intercept	5.59	< 0.01
Baseline	-2.22	< 0.01
Infocost=high	-1.58	< 0.01
1/Period	-11.83	< 0.01
Baseline*1/Period	21.96	< 0.01
Infocost=high*1/Period	30.38	< 0.01
Baseline*Infocost=high*1/Period	-26.72	< 0.01
Random Effects: Intercept at the group level = 0.84		
Intercept at the individual level = 0.76		
Number of observations: 5760		
Number of groups: 48 / Number of individuals: 144		

Table 8: Results of the statistical analysis of efforts for the two mechanisms.

treatments (4.29 units in BaseLow and 4.63 units in BaseHigh). Instead, the estimated initial effort level in CommitHigh is significantly higher (5.7 units). Starting out from these different initial levels, the estimation confirms different trends for the two commitment treatments. The estimated average effort in CommitHigh decreases to 4.38 units in the last period, while average effort provision in CommitLow flourishes and increases to 5.37 units in the last period. Irrespective of the commitment treatment, both estimated average effort levels for the final period are significantly higher than the analogous estimations for the baseline (3.57 in BaseLow and 3.65 in BaseHigh). The significant impact of the variable $1/Period$ together with its significant interaction effects underlines that the estimated average effort level is increasing at a decreasing marginal rate in CommitLow. In fact, the estimated level is predicted to asymptotically approach an average of 5.5 units from below. Instead, in CommitHigh the estimated average effort level is decreasing at a decelerating marginal rate, converging to the equilibrium level of 4 units from above. Concluding, the regression predicts that the pairwise commitment mechanism induces average effort levels which are at least as high as the equilibrium effort level induced by the formation of pairwise stable networks. As this is significantly higher than the equilibrium level of the effort game under the voluntary contribution mechanism (to which average effort levels converge in the baseline treatments) the pairwise commitment mechanism proves to be significantly effective. In fact, its actual efficacy depends on the cost of monitoring. When this cost is low, the pairwise commitment mechanism, although falling short of inducing full cooperation/efficiency, increases and afterwards stabilizes cooperation at a level well above the equilibrium.

It is now of immediate interest to see to what extent and in which way subjects made use of the possibility to form pairwise cooperation commitments. A first rough measure is the average network density, which captures the average number of cooperation commitments $l(g)$ in the observed cooperative networks. Table 9 summarizes its evolution over the course of the session.

Irrespective of the treatment, we find that pairs of subjects do immediately after the introduction of the mechanism make use of the possibility to form cooperation commitments. On average, 1.58 cooperation commitments are formed per group right at the end of the 10th period. This falls

	Period 11 to 20	Period 21 to 30	Period 31 to 40	Period 41 to 50	All periods
CommitLow	1.58	1.58	2.08	2.33	1.9
CommitHigh	1.58	1.75	1.67	1.58	1.65

Table 9: Evolution of average network density.

only slightly short of the density of the predicted pairwise stable cooperative network, viz., 2. In fact, the observed increase in the aggregate average effort level in CommitLow is backed up by a continuous increase in network density over the course of the 50 periods. On the other hand, network density decreases in CommitHigh starting with $k = 30$, which also reflects the trend in average effort provision in this treatment. In CommitLow the average network density moves beyond the equilibrium density of two cooperation commitments per group, while it consistently stays below that level in CommitHigh. Interestingly, the data indicate that the observed average effort levels are not fully accounted for by the number of formed cooperation commitments. Instead, subjects move beyond the effort they committed to, as the average density of 1.9 cooperation commitments in CommitLow translates into an average (individual) effort provision of 3.8 units while we observe an average of 4.87 units. Analogously, the density of 1.65 cooperation commitments observed in CommitHigh induces an average effort of 3.3 units, while we observe an average effort provision of 4.71 units.

In order to see, which networks support cooperation beyond the equilibrium level, and to what extent incentive-compatible networks have been formed at all, table 10 summarizes the relative frequencies of the different observed *non-empty* networks together with the network contingent effort levels. We distinguish between the different positions (defined by the number of 'links', i.e., cooperation commitments) in asymmetric networks. As cooperation commitments - once finalized - where exogenously enforced in our experimental implementation, subjects could only deviate by providing a higher effort than the one they committed to. We quote deviations in relative terms, i.e., as fractions of the effort level that subjects committed to given their position in the network ('na' denotes those network positions in which subjects did not commit themselves).

We find that, irrespective of the treatment, the largest fraction of observed networks is indeed pairwise stable (g^2). Nevertheless, pairwise stable networks are formed more frequently under low than under high monitoring costs. Moreover, while 10% of all finalized networks in CommitLow are efficient, we do not observe any efficient network in the high monitoring cost treatment. Now looking at the observed network-contingent effort levels, the most interesting finding is that especially pairwise stable networks induce effort provision levels considerably beyond the level that their formation induces. Secondly, in asymmetric networks, we find some support for the hypothesis that subjects take not only their own cooperation commitments but instead the whole structure of the cooperative network into account. More precisely, subjects that have formed relatively fewer commitments than others, deviate relatively more from the effort they have committed to, which is washing out some of the asymmetry manifested by the network. But compensation is far from

Network	Links	Relative NW-frequency: Average effort (effort's std.; # obs.) - Relative deviation *CommitLow*	*CommitHigh*
g°	2	0.10: 8.03 (0.17; 200) - 0.38%	—
g^2	1	0.52: 4.86 (1.03; 1000) - 21.40%	0.42: 4.99 (1.35; 800) - 24.81%
$g^{\star+1}$	3	—	0.02: 12.00 (0.00; 10) - 0.00%
	2	—	0.02: 8.30 (0.66; 20) - 3.75%
	1	—	0.02: 4.00 (0.00; 10) - 0.00%
g^l	2	0.06: 8.08 (0.38; 60) - 1.04%	0.08: 8.51 (0.89; 80) - 6.41%
	1	0.06: 4.42 (0.79; 60) - 10.42%	0.08: 5.31 (1.50; 80) - 32.81%
$g^{\triangle}$	2	—	0.02: 8.07 (0.37; 30) - 0.83%
	0	—	0.02: 0.00 (0.00; 10) - na
$g^{2\star}$	2	0.08: 8.33 (0.83; 40) - 4.06%	0.10: 8.02 (0.14; 50) - 0.25%
	1	0.08: 4.63 (1.25; 80) - 15.63%	0.10: 4.60 (0.99; 100) - 15.00%
	0	0.08: 2.03 (1.91; 40) - na	0.10: 2.66 (1.76; 50) - na
g^1	1	0.08: 4.40 (0.89; 80) - 10.00%	0.21: 4.83 (1.21; 200) - 20.63%
	0	0.08: 2.51 (1.81; 80) - na	0.21: 2.64 (1.67; 200) - na

Table 10: Observed network-contingent average efforts and relative deviation from committed effort.

complete, meaning that the actual effort provision levels in asymmetric networks remain unequal between subjects in different positions.

To pin down differences between the two treatments and to identify dynamics we once more recur to a mixed effects linear regression, now estimating the average effort as a function of the formed network. We capture the different positions in different networks by including both the number of own links (*Own_links*) as well as the degree of network asymmetry (*Asym*).[32] Additionally, we include *Period* and *Infocost* and all possible interaction effects in the original complete model.[33] The results - after appropriate reduction of the model[34] - are summarized in table 11.

Considering only the qualitative results, the regression suggests that in the pairwise stable network, the initial average effort in Period 11 is higher in CommitLow compared to CommitHigh. But while it remains stable in CommitLow, it increases in CommitHigh, moving beyond the level estimated for CommitLow. Whenever the formed network is weakly asymmetric (Asym = 1, viz. in g^1 and g^l), then the estimated average efforts are higher in CommitHigh than in CommitLow. In intermediately asymmetric structures such as $g^{2\star}$ the cost of monitoring has no significant impact, but we observe a significant decrease in the estimated effort level from the first (Period 11) to the last period, irrespective of player's position in the network. Concluding, we observe that coopera-

[32] The measure *Asym* is derived from the variance in the number of links across different players in a network. Concretely, it is given by $Asym(g) = [4\sum_{i=1}^{4} l_i(g) - (\sum_{i=1}^{4} l_i)^2]/4$. Accordingly, the measure is taking the value 0 in all symmetric networks ($g^{\emptyset}$, g^2, g°, g^N), the value 1 in weakly asymmetric structures (g^1, g^l and g^5), the value 2 in intermediately asymmetric structures ($g^{2\star}$, $g^{\star+1}$), and the value 3 in very asymmetric structures ($g^{\star}$ and $g^{\triangle}$). Given the two variables 'Own links' and 'Asym', we can unambiguously distinguish between all positions in all networks, except for players with 1 link in g^1 and g^l and players with 2 links in $g^{2\star}$ and in $g^{\star+1}$.

[33] We exclude the observations for $g^{\star+1}$ and $g^{\triangle}$ as these two networks have only been formed by one group in one block, which in turn implies that all effort levels for the position with one link in $g^{\star+1}$ and for the singleton in $g^{\triangle}$ have been generated by the same subject.

[34] We include random intercepts at the individual level. Random intercepts at the group level do not improve the fit. We controlled for autocorrelation and heteroscedasticity of the residuals.

	Value	p-value
Intercept	4.06	< 0.01
1 link	1.30	< 0.01
2 links	4.42	< 0.01
1 link * Infocost = high	-1.15	< 0.01
Asym = 2 * Period	-0.05	< 0.01
Asym = 1 * Infocost = high	1.24	< 0.01
1 link * Period * Infocost	0.04	< 0.01
Random Effects: Intercept = 0.71		
Number of observations: 3760		
Number of individuals: 96		

Table 11: Results of the statistical analysis of network-contingent efforts.

tion beyond the equilibrium level in the pairwise stable network is immediate in CommitLow, while subjects learn to 'mutually cooperate' in CommitHigh, finally moving beyond the level observed under low monitoring costs. Also we observe that subjects obviously cope better with weakly asymmetric networks in the CommitHigh treatment, which is in fact where most asymmetric networks are observed.

The up to now presented aggregates again mask a considerable heterogeneity between different groups, which is even more pronounced than in the first experiment. First confirming the aggregate picture in CommitLow, we find that the mass of groups has shifted towards a higher overall cooperation level compared to BaseLow (see, table 16 in the Appendix). But even though all groups are providing effort levels beyond the equilibrium level induced by the formation of pairwise stable networks, groups under CommitLow exhibit a dispersion of overall average cooperation rates between 36.4% and 75.9%, where 1/4 of the groups manage to achieve full cooperation, by converging to the efficient average effort level of 8 units towards the end of the session. And while the large majority of groups (5/6) succeeds in either stabilizing (these are 1/3 of all groups, viz. Groups 1, 2, 9, and 12) or even in considerably increasing (these are 1/2 of all groups, viz. Groups 3-5, 7, 8, and 10) their rate of cooperation over the course of the 50 periods, there remain two groups which exhibit a slightly decreasing trend (Groups 6, and 11). This shows that some groups succeed in using the commitment mechanism (which does not per se implement the efficient outcome) to *completely solve* the social dilemma, while cooperation nevertheless deteriorates in others.

Also in CommitHigh the mass of groups is shifted towards a higher overall average effort level, compared to BaseHigh (see, table 17 in the Appendix). But different from CommitLow the lowest performing groups are now providing an overall average effort level *below* the equilibrium level induced by the formation of pairwise stable networks. Effectively, 42% of the groups exhibit a cooperation level around the equilibrium level, viz. a cooperation rate below 40% (Groups 1, 4, 8, 10, and 11), while this is only the case for 25% of the groups in CommitLow. On the other hand, underlining the enormous heterogeneity across different groups, one group succeeds in using the mechanism to converge to full cooperation. Concerning the observed dynamics of the average effort levels, in CommitHigh only half of the groups are able to stabilize (Groups 4, 5, and 11) or increase (Groups 1, 3, and 9) their cooperation rate, while cooperation rates deteriorate in the remaining other half of the groups (Groups 2, 6-8, 10, and 12).

We now put these results into perspective, by looking precisely at the emergence of cooperative networks in the two commitment treatments as summarized in table 12.

	Network	Period 11 to 20	Period 21 to 30	Period 31 to 40	Period 41 to 50
	pw-stable	41.67%	58.33%	58.33%	50.00%
CommitLow	efficient	0.00%	8.33%	8.33%	25.00%
	empty	16.67%	25.00%	8.33%	8.33%
	others (asym.)	41.67%	8.33%	25.00%	16.67%
	pw-stable	8.33%	66.67%	41.67%	50.00%
CommitHigh	efficient	0.00%	0.00%	0.00%	0.00%
	empty	33.33%	0.00%	8.33%	16.67%
	others (asym.)	58.33%	33.33%	50.00%	33.33%

Table 12: Relative fraction and emergence of cooperative networks.

The initial fraction of groups in CommitLow that form a pairwise stable network right after the introduction of the commitment mechanism is with 41.67% comparatively high. After that the fraction of groups forming pairwise stable networks increases, before the fraction decreases, largely because pairwise stable networks are replaced by efficient networks. At the end of period 50, 3/4 of the groups in CommitLow have either converged to a pairwise stable (50%) or efficient (25%) network. The precise network evolution paths in different groups (provided in table 18 in the Appendix) underlines that groups differ largely with respect to their ability to make use of the pairwise commitment mechanism. Whereas three of the twelve groups exhibit a very systematic cooperative network evolution path, in which they move from a pairwise stable network to an efficient network (Group 4 does so very fast), other groups (1/4, viz. Groups 6, 7 and 9) either cannot agree on the formation of any pairwise commitments at all (Group 9) or do not reach or sustain pairwise stable networks.

In total, 42% of all observed networks are pairwise stable in CommitHigh which is slightly less than in CommitLow. But the more dramatic difference between the two treatments concerns the initial fraction of pairwise stable networks. While almost half of the groups in CommitLow managed to form a pairwise stable network immediately after the commitment mechanism had been introduced, this is only the case for about 8% of the groups in CommitHigh. After this initial low start, the fraction of pairwise stable networks increases in CommitHigh, with 1/2 of the groups having converged to a pairwise stable network at the end of period 40. But different from CommitLow, none of the groups manages to form an efficient network. Overall, the evolution of networks in the different groups is far less structured than in CommitLow (see, table 19 in the Appendix). Moreover, a considerable and almost constant number of the networks observed in CommitHigh is asymmetric.

Finally, we take a look at the monitoring activity. In CommitLow subjects on average over the 50 periods monitored 9.13 effort choices of their partners. In fact, this is less than half of the information that subjects bought in BaseLow.[35] But different from BaseLow, the monitoring activity is fully concentrated on the initial periods, where it is much higher than in BaseLow. Consequently, even though subjects did not monitor their partners' behavior at a high level over the complete course of the session, some subjects very intensively monitored their partners at the beginning of the session. Again, and as intended, the monitoring activity has been much reduced by the introduction of considerable monitoring costs in CommitHigh. Here subjects bought an average of 4.71 information pieces which is slightly more than in BaseHigh.

4.4 Discussion

The commitment mechanism is effective in inducing a higher aggregate average effort level than the voluntary contribution mechanism. But more than that, the mechanism helps cooperation *beyond* the equilibrium level induced by the formation of pairwise stable networks. When groups form a pairwise stable network, they tend to cooperate beyond the committed level, while some groups even manage to form efficient networks. We observe largely heterogeneous behavior between different groups. While some groups succeed in using the mechanism to support full cooperation, cooperation in other groups decreases and some groups completely refrain from forming any commitments.

Interestingly, the performance of the mechanism depends on the cost of individual monitoring. Individual monitoring, when available at a low cost, increases the efficacy of the pairwise commitment mechanism, allowing a larger fraction of groups to not only stabilize but instead to increase aggregate cooperation rates well beyond the equilibrium level over time. In fact, monitoring seems to be particularly important in the initial phase of groups' interactive relationships. Here individualized monitoring seems to help subjects to coordinate on a 'cooperative norm' which facilitates the immediate formation of pairwise stable networks, which - given the observed beyond-equilibrium cooperation in pairwise stable networks - seems to later on facilitate the coordination on efficient networks.

5 Cooperative networks and social preferences: a theoretical explanation of the experimental evidence

In essence, our experimental results can be condensed into the following stylized findings: 1) both under the voluntary contribution and the pairwise commitment mechanism we observe that while some groups provide effort levels considerably above the respective equilibrium others are providing efforts just at or even below the equilibrium, 2) the cost of monitoring has no systematic impact on the distribution of groups into 'beyond-equilibrium' cooperating and 'at-or-below equilibrium' cooperating groups under the voluntary contribution mechanism, whereas 3) under the pairwise commitment mechanism, high monitoring costs quarter the fraction of highly cooperating groups while at the same time they increase the fraction of groups that are cooperating below the

[35] Individuals are again heterogeneous with respect to their monitoring activity: 35.4% of all subjects never buy any information, while 70.8% buy less than 10 pieces of information.

equilibrium level five-fold.[36]

These findings directly relate to the well-established experimental evidence collected in diverse other strategic environments which suggests that there are individuals that do not solely care about their own material well-being. Several modeling approaches have been put forward to formally account for this evidence.[37] Their common basis is the introduction of social preferences while retaining the rationality assumption. Their difference lies in the way they conceptualize social preferences. In a seminal paper by Fehr and Schmidt (1999), social preferences take the form of inequity aversion. It is important to note that inequity aversion plays the role of a 'black box', which in Fehr and Schmidt's operationalization has proven very effective in capturing the interactive *outcomes* in several different contexts but which does not sufficiently reflect the *forces* at play. There are approaches that have tried to model the latter by, e.g., formalizing intention-based reciprocity (Rabin, 1993; Charness and Rabin, 2002; Dufwenberg and Kirchsteiger, 2004; Falk and Fischbacher, forthcoming). But the gained realism comes at the expense of a loss in these models' tractability which is a direct result of their dependence on psychological game theory. In the light of this trade-off we will in the following rely on Fehr and Schmidt's (1999) model and investigate whether inequity aversion may provide a *tentative* explanation of our stylized findings.

According to the results of our first experiment, repeated game effects cannot be considered as the main driving force of the observed behavior. Therefore, and for the sake of simplicity, we leave out repeated game considerations in our extended theoretical setting and we compare its predictions to the behavior of subjects in the late periods of interaction. As an additional restriction, we only consider the parameterized version of the effort game which has been implemented in the laboratory.

5.1 The effort game with inequity averse players

Four players simultaneously exert some efforts the benefits of which accrue equally to all players. As in the second section of the paper, e_i denotes player i's effort level, where $e_i \in \{0, 1, \ldots, 12\}$ and we assume that the level of effort exerted involves a cost of $\frac{1}{4}e_i^2$ which is private to player $i \in N = \{1, 2, 3, 4\}$. In line with Fehr and Schmidt (1999) we assume that player i's utility function can be written as

$$U_i(e_i, e_{-i}) = \pi_i(e_i, e_{-i}) - \frac{\alpha_i}{3} \sum_{j \neq i} max\{\pi_j - \pi_i, 0\} - \frac{\beta_i}{3} \sum_{j \neq i} max\{\pi_i - \pi_j, 0\}$$

where $\pi_i(e_i, e_{-i}) = \sum_{i=1}^{4} e_i - \frac{1}{4}e_i^2$ is player i's material payoff, and $\alpha_i \geq \beta_i \geq 0$, and $\beta_i < 1$. Consequently, if $\alpha_i = \beta_i = 0$, player i is a selfish player. Otherwise, player i is inequity averse. The model assumes that an inequity averse player experiences inequity both when he is comparatively better off and when he is comparatively worse off than other players, but that his disutility from being comparatively advantaged is smaller (α_i) than from being comparatively disadvantaged (β_i).

Fehr and Schmidt's approach naturally lead to an incomplete information setting in which the

[36]When considering the last 5 periods: Under the voluntary contribution mechanism, irrespective of the treatment, 9 groups (among 18) are cooperating just slightly above the equilibrium of 2 units, while 3 groups cooperates at a high level of more than 4 units. In CommitLow 4 groups are cooperating efficiently at or beyond 6 units, while one group is cooperating at 3 units, viz., below the equilibrium. Instead in CommitHigh, one group is cooperating beyond 6 units, while 5 groups are cooperating below 4 units and one group is even contributing below 2 units.

[37]For surveys, see Sobel (2004), and Fehr and Schmidt (2003).

couple (α_i, β_i) is private information to player i who only knows the distribution of the other players' inequity parameters (this distribution is in fact assumed to be commonly known). In the following we however derive the predictions of the effort game with inequity averse players under the assumption that the social preferences are common knowledge. The reason is twofold: first, we conjecture that such a complete information setting approximates our experimental setup with a low monitoring cost where subjects have collected a reasonable amount of evidence about their partners' behavior, and second this will enable us to informally discuss the predictions in the incomplete information setting which approximates our experimental setup with a high monitoring cost where almost no evidence about the partners' behavior has been collected. Needless to say, the claim that by collecting some pieces of information about their partners' behavior subjects have been able to figure out those partners' preferences is strong.[38]

5.2 Predicted behavior with a low monitoring cost: the complete information setting

In the following, we derive the predictions of the effort game with inequity averse players both under the voluntary contribution mechanism and the pairwise commitment mechanism. These predictions are an attempt to account for the facts that groups are heterogeneous and that the efficient outcome can be observed under the pairwise commitment mechanism but not under the voluntary contribution mechanism. Below, we provide two propositions summarizing the predictions of the effort game with inequity averse players, one per mechanism. However, as our extended theoretical setting leads to numerous predictions depending on the possible values of the inequity parameters, we only list the equilibria which can be 'reasonably' expected. Indeed, Fehr and Schmidt have provided an estimation of the realistic values of the αs and βs (see, Fehr and Schmidt (1999), Table III, p. 844): accordingly, α cannot be expected to be greater than 4, while β cannot be expected to be greater than 3/5. The complete set of equilibria is provided in the Appendix.

Proposition 4 *Under the voluntary contribution mechanism, if $\alpha_i < 1/3$ and $\beta_i < 1/5$, player i has a (strictly) dominant strategy to exert effort $e_i = e^* = 2$, and the predictions of the effort game played by inequity averse players with commonly known preferences are given by:*

- *If $\alpha_i \geq 1/3$ or $\beta_i \geq 1/5$, but $\alpha_j < 1/3$ and $\beta_j < 1/5$, $\forall j \neq i$, $i, j \in N$, $(2, 2, 2, 2)$ is the unique equilibrium.*
- *If $\alpha_i \geq 1/3$ or $\beta_i \geq 1/5$, $\forall i \in \{1, 2\}$, but $\alpha_j < 1/3$ and $\beta_j < 1/5$, $\forall j \in \{3, 4\}$, in addition to the above mentioned equilibrium, $(3, 3, 2, 2)$ and $(1, 1, 2, 2)$ are equilibria where respectively $\beta_i - 2\alpha_i \geq 3/5$ and $\alpha_i - 2\beta_i \geq 1$.*
- *If $\alpha_i \geq 1/3$ or $\beta_i \geq 1/5$, $\forall i \in \{1, 2, 3\}$, but $\alpha_4 < 1/3$ and $\beta_4 < 1/5$, in addition to the above mentioned equilibria, $(3, 3, 3, 2)$ and $(1, 1, 1, 2)$ are equilibria where respectively $2\beta_i - \alpha_i \geq 3/5$ and $2\alpha_i - \beta_i \geq 1$.*

[38] Moreover, the fact that a subject monitors his partners is far from being enough for the preferences to become common knowledge. Still, we consider reasonable to assume that a subject who monitors his partners also believes that those partners monitor him, and additionally believes that his partners expect him to monitor them. Consequently, subjects' preferences can become mutual knowledge at a sufficiently high order.

- *If* $\alpha_i \geq \frac{1}{3}$ *or* $\beta_i \geq \frac{1}{5}$, $\forall i \in \{1,2,3,4\}$, *in addition to the above mentioned equilibria,* (e,e,e,e) *with* $e \in \{3,4,5\}$ *is an equilibrium if* $\beta_i \geq (2e-5)/(2e-1)$, $(1,1,1,1)$ *is an equilibrium if* $\alpha_i \geq 1$, $(2,1,0,0)$ *is an equilibrium if* $\alpha_1 \leq 3/31$, $1 \leq 2\alpha_2 - \beta_2 \leq 9$ *and* $4\alpha_j - 2\beta_j \geq 9$, $\forall j \in \{3,4\}$.

In an extension of Fehr and Schmidt's (1999) terminology we refer to a player that has a strictly dominant strategy as a selfish player and to all other players as fair players. Proposition 4 shows that - when assuming realistic preferences - already with two fair players it is possible to sustain both an equilibrium modestly beyond the standard equilibrium and one modestly below the latter, where basically the same holds for three fair players. With four fair players it is possible to sustain quite considerable effort levels (at 5 units), but also to obtain efforts at a below-equilibrium level of 1 units per player.

In conclusion, the introduction of inequity-averse players into the implemented version of the effort game without a possibility to form pairwise commitments indeed allows us to reproduce our first stylized fact. The model explains both cooperation 'beyond standard equilibrium' as well as 'at-or-below standard equilibrium' and even correctly predicts the levels that we observe.

Our next proposition establishes that the efficient network g° is incentive-compatible only in groups composed of four fair players, and that it can be the only pairwise stable network comprising at least three pairwise commitments.

Proposition 5 *In the effort game with the possibility to form pairwise commitments,*

- *If* $\alpha_i < \frac{1}{3}$ *and* $\beta_i < \frac{1}{5}$, $\forall i \in \{1,2,3,4\}$, *the unique pairwise stable network is* g^2.

- *If* $\alpha_1 \geq \frac{1}{3}$ *or* $\beta_1 \geq \frac{1}{5}$, *but* $\alpha_j < \frac{1}{3}$ *and* $\beta_j < \frac{1}{5}$, $\forall j \in \{2,3,4\}$, *the unique pairwise stable network is* g^2.

- *If* $\alpha_i \geq \frac{1}{3}$ *or* $\beta_i \geq \frac{1}{5}$, $\forall i \in \{1,2\}$, *but* $\alpha_j < \frac{1}{3}$ *and* $\beta_j < \frac{1}{5}$, $\forall j \in \{3,4\}$, *the unique pairwise stable network is* g^2.

- *If* $\alpha_i \geq \frac{1}{3}$ *or* $\beta_i \geq \frac{1}{5}$, $\forall i \in \{1,2,3\}$ *but* $\alpha_4 < \frac{1}{3}$ *and* $\beta_4 < \frac{1}{5}$, *no network with* $l(g) \geq 3$ *is pairwise stable.*

- *If* $\alpha_i \geq \frac{1}{3}$ *or* $\beta_i \geq \frac{1}{5}$, $\forall i \in \{1,2,3,4\}$, *among all networks with* $l(g) \geq 3$ *only* g° *is pairwise stable if* $\frac{1}{2} \leq \beta_i \leq \frac{5}{9}$, *and* $\alpha_i > 0$, $\forall i \in \{1,2,3,4\}$.

Again, this prediction is in line with our experimental evidence. Indeed, we observe a considerable number of g^2 networks, which can be sustained with groups composed of none, one, two, three and four inequity averse players, but also a small fraction of g° networks, which need the presence of four quite extreme inequity averse players.

In the following, we provide an informal discussion about the expected impact of a high monitoring cost in both mechanisms.

5.3 Predicted behavior with a high monitoring cost: the incomplete information setting

Fehr and Schmidt (1999) have been able to explain the observed behavior in many experimental games according to the following two discrete distributions of inequity averse parameters: $\alpha = 0$ in 30 percent of the cases, $\alpha = 1/2$ in 30 percent of the cases, $\alpha = 1$ in 30 percent of the cases, $\alpha = 4$ in 10 percent of the cases, $\beta = 0$ in 30 percent of the cases, $\beta = 1/4$ in 30 percent of the cases, and $\beta = 0.6$ in 40 percent of the cases. Under the assumption that this distribution of preferences is commonly known to the players, it seems reasonable to conclude that in an incomplete information version of the effort game with the possibility to form pairwise commitments no network other than g^2 is pairwise stable. Indeed, as soon as one selfish player is part of the group, no network with at least three commitments is pairwise stable, and this event is quite likely. Therefore, a low cost of monitoring is more favorable to the emergence of efficient cooperative networks. On the other hand, in an incomplete information version of the effort game under the voluntary contribution mechanism, the expected equilibrium effort is some kind of an average of the equilibrium efforts exerted in the complete version of the game which suggests that the cost of monitoring should not have a significant impact.

6 Conclusion

In this paper we evaluated the efficacy of voluntary pairwise commitment formation to solve a collective action problem. Our results suggest that it may not in general be necessary for the legislative authority to centrally impose mechanisms which directly implement mutual cooperation. Instead, decentralized negotiation provides a social mechanism which is effective in generating and sustaining significantly higher rates of cooperation than the voluntary contribution mechanism.[39] But the efficacy of the mechanism in this respect depends on the informational conditions. While under high monitoring costs, the aggregate cooperation rate converges to the equilibrium rate over time, it increases and finally stabilizes at a level well beyond the equilibrium when monitoring costs are low. One quarter of the groups in the low monitoring cost treatment were able to achieve *full cooperation* in the later periods of the game by coordinating on the formation of efficient networks. This contrasts starkly with the high monitoring cost treatment in which none of the groups managed to coordinate on efficient networks. In general, we observe a large heterogeneity concerning the capacity of different experimental groups to make use of the commitment mechanism to support cooperation. This replicates the empirical fact, that only some communities in the field successfully resolve collective action problems by relying on informal organizational modes such as cooperative agreements between neighbors, while others fail in this respect.

Our stylized findings, viz., 1) the large heterogeneity between different groups under both the

[39] In fact, the pairwise commitment mechanism does well also in comparison to other mechanisms that have been tested in the laboratory: although most punishment mechanisms have been shown to be effective in inducing high contribution levels they are - due to the cost of punishment - not efficiency-enhancing (Fehr and Gächter, 2000, 2002; Bochet, Page, and Putterman, forthcoming). Those that are, are not informal but require an intervening regulator to run the mechanism as, e.g., in the tax-subsidy mechanism implemented by Falkinger, Fehr, Gächter, and Winter-Ebmer (2000). Instead, the pairwise commitment mechanism may just require the central authority to uphold contract law but not to administer the mechanism. Other in the literature considered mechanisms have been highly artificial (Kurzban, McCabe, Smith, and Wilson, 2001; Goren, Kurzban, and Rapoport, 2003).

voluntary contribution and the pairwise commitment mechanisms, 2) the formation of efficient cooperative networks by some groups, and 3) the differential impact of the cost of monitoring under the two mechanisms can be explained by inequity-aversion. Fehr and Schmidt's (1999) model effectively captures our data and underlines that fairness concerns drive the emergence of cooperative networks and can explain the network-contingent cooperation rates in the considered context.

Appendix

Descriptive statistics of the first experiment

	Average effort (effort's standard deviation):		Cooperation rate
	First 10 periods	Last 10 periods	All periods
Group1	4.15 (0.84): 35.8%	4.38 (0.13): 39.6%	3.97 (0.63): 32.8%
Group2	3.53 (0.95): 25.4%	4.25 (0.66): 37.5%	3.89 (0.85): 31.5%
Group3	2.83 (0.79): 13.8%	2.65 (0.46): 10.8%	2.68 (0.70): 11.3%
Group4	4.35 (1.21): 39.2%	4.3 (0.76) : 38.3%	4.25 (0.85): 37.5%
Group5	5.15 (0.73): 52.5%	2.3 (0.84) : 5%	4.07 (1.33): 34.4%
Group6	4.13 (1.06): 35.4%	3.05 (0.60): 17.5%	3.55 (0.63): 25.8%

Table 13: Groups in NoFeedback.

	Average effort (effort's standard deviation):		Cooperation rate
	First 10 periods	Last 10 periods	All periods
Group 1	6.15 (1.49): 69.2%	3.83 (0.81): 30.4%	4.6 (1.27): 43.3%
Group 2	3.40 (0.87): 23.3%	3.53 (0.84): 25.4%	3.34 (1.02): 22.3%
Group 3	3.25 (0.63): 20.8%	2.25 (0.44): 4.2%	2.84 (0.72): 13.9%
Group 4	3.85 (0.64): 30.8%	2.65 (0.47): 10.8%	3.16 (0.76): 19.3%
Group 5	6.13 (0.27): 68.8%	5.6 (0.8): 60%	5.93 (0.63): 65.4%
Group 6	4.58 (1.53): 42.9%	3.4 (1.59) : 23.3%	3.67 (1.54): 27.8%

Table 14: Groups in BaseLow.

	Average effort (effort's standard deviation): Cooperation rate		
	First 10 periods	Last 10 periods	All periods
Group 1	5.95 (1.39): 65.8%	3.58 (0.80): 26.3%	4.44 (1.28): 40.6%
Group 2	6.03 (1.00): 67.1%	4.63 (0.98): 43.8%	4.79 (1.11): 46.5%
Group 3	2.65 (0.60): 10.8%	2.25 (0.53): 4.2%	2.43 (0.60): 7.2%
Group 4	5.38 (0.95): 56.3%	3.88 (0.41): 31.3%	4.42 (0.93): 40.3%
Group 5	4.03 (1.02): 33.8%	3.25 (0.61): 20.8%	3.21 (0.94): 20.2%
Group 6	4.03 (1.13): 33.8%	2.43 (0.65): 7.1%	2.99 (1.20): 16.5%

Table 15: Groups in BaseHigh.

Descriptive statistics of the second experiment

	Average effort (effort's standard deviation): Cooperation rate		
	First 10 periods	Last 10 periods	All periods
Group 1	4.20 (0.86): 36.7%	4.68 (0.29): 44.6%	4.73 (0.93): 45.4%
Group 2	5.18 (1.25): 52.9%	4.73 (0.46): 45.4%	4.68 (1.30): 44.7%
Group 3	2.80 (0.75): 13.3%	5.23 (0.25): 53.8%	4.19 (1.00): 36.4%
Group 4	3.78 (0.28): 29.6%	8.00 (0.00): 100.0%	6.56 (1.83): 75.9%
Group 5	2.55 (1.03): 9.2%	8.00 (0.00): 100.0%	4.58 (1.89): 42.9%
Group 6	3.75 (0.77): 29.2%	2.95 (0.28): 15.8%	4.24 (1.09): 37.3%
Group 7	2.88 (1.08): 14.6%	6.10 (0.17): 68.3%	4.41 (1.31): 40.2%
Group 8	4.25 (0.75): 37.5%	8.15 (0.17): 102.5%	5.09 (1.59): 51.4%
Group 9	4.28 (0.95): 37.9%	4.73 (0.89): 45.4%	4.67 (1.24): 44.5%
Group 10	3.08 (1.10): 17.9%	4.53 (0.25): 42.1%	4.21 (0.97): 36.8%
Group 11	5.38 (0.64): 56.3%	4.48 (0.61): 41.3%	5.52 (0.87): 58.7%
Group 12	5.50 (0.24): 58.3%	5.50 (0.37): 58.3%	5.59 (0.55): 59.8%

Table 16: Groups in CommitLow.

	Average effort (effort's standard deviation): Cooperation rate		
	First 10 periods	Last 10 periods	All periods
Group 1	3.20 (0.63): 20.0%	4.38 (0.21): 39.6%	3.71 (0.65): 28.5%
Group 2	5.10 (0.77): 51.7%	2.73 (0.89): 12.1%	4.52 (1.10): 41.9%
Group 3	3.48 (1.11): 24.6%	6.30 (1.38): 71.7%	5.21 (1.50): 53.4%
Group 4	4.23 (0.45): 37.1%	3.28 (0.36): 21.3%	4.16 (0.57): 35.9%
Group 5	4.03 (1.06): 33.8%	4.65 (0.52): 44.2%	4.61 (0.82): 43.4%
Group 6	6.85 (0.89): 80.8%	5.20 (1.08): 53.3%	5.94 (1.21): 65.7%
Group 7	5.35 (1.07): 55.8%	5.68 (0.70): 61.3%	5.24 (1.46): 54.0%
Group 8	4.55 (1.81): 42.5%	2.35 (0.61): 5.8%	4.00 (1.40): 33.3%
Group 9	4.58 (1.01): 42.9%	8.15 (0.21): 102.5%	6.20 (1.42): 70.0%
Group 10	4.18 (2.01): 36.3%	2.60 (0.52): 10.0%	4.30 (1.45): 38.3%
Group 11	3.70 (0.79): 28.3%	3.43 (0.47): 23.8%	3.61 (0.59): 26.8%
Group 12	5.60 (1.20): 60.0%	4.70 (0.51): 45.0%	5.03 (0.83): 50.5%

Table 17: Groups in CommitHigh.

	Periods 11-20	Periods 21-30	Periods 31-40	Periods 41-50
Group 1	g^l	$g^\emptyset$	g^2	g^2
Group 2	$g^\emptyset$	$g^\emptyset$	g^2	g^2
Group 3	g^1	g^2	g^2	g^2
Group 4	g^2	g°	g°	g°
Group 5	g^2	g^2	g^2	g°
Group 6	$g^{2\star}$	g^2	$g^{2\star}$	g^1
Group 7	g^1	g^2	$g^{2\star}$	g^l
Group 8	g^2	g^2	g^2	g°
Group 9	$g^\emptyset$	$g^\emptyset$	$g^\emptyset$	$g^\emptyset$
Group 10	$g^{2\star}$	g^1	g^2	g^2
Group 11	g^2	g^2	g^l	g^2
Group 12	g^2	g^2	g^2	g^2

Table 18: Evolution of networks in CommitLow.

	Periods 11-20	Periods 21-30	Periods 31-40	Periods 41-50
Group 1	$g^\emptyset$	g^1	g^1	g^2
Group 2	$g^{2\star}$	g^2	$g^{2\star}$	$g^\emptyset$
Group 3	g^l	g^2	g^2	g^2
Group 4	$g^\emptyset$	g^1	g^2	g^1
Group 5	$g^{2\star}$	g^2	g^2	g^2
Group 6	g^l	g^2	g^2	g^2
Group 7	g^l	g^2	$g^\emptyset$	g^2
Group 8	$g^\emptyset$	$g^{2\star}$	g^1	$g^\emptyset$
Group 9	g^2	g^2	g^l	$g^{\star+1}$
Group 10	$g^\triangle$	g^2	$g^{2\star}$	g^1
Group 11	$g^\emptyset$	g^1	g^1	g^1
Group 12	g^1	g^2	g^2	g^2

Table 19: Evolution of networks in CommitHigh.

Proofs of Section 2

Proof of Proposition 1. Consider the two cooperative networks g' and g which are elements of $G = \{g \mid g \subseteq g^N\}$ such that $g' = g \cup ij$ $(g \neq g^N)$. If $l_i(g) = l_j(g) = 0$ then $\pi_i^*(g') - \pi_i^*(g) = \pi_j^*(g') - \pi_j^*(g) = 2\,(e_{BA} - e^*) - c\,(e_{BA}) + c\,(e^*) > 0$ as $2 = c'\,(e_{BA}) > (c\,(e_{BA}) - c\,(e^*))\,/\,(e_{BA} - e^*)$. In other words, at most one player does not form a cooperation commitment in a pairwise stable cooperative network. We now prove that no player agrees to form strictly more than one cooperation commitment in a pairwise stable cooperative network. Indeed, if $l_i(g') \geq 2$ and $l_j(g') \geq 2$ then $\pi_i^*(g) - \pi_i^*(g') = c\,(e_{BA}\,l_i(g')) - c\,(e_{BA}\,l_i(g)) - 2e_{BA} > 0$ as $c'\,(e_{BA}) = 2$ and the effort cost function is strictly convex. For similar reasons, if $l_i(g') \geq 2$ and $l_j(g') = 1$ then $\pi_i^*(g) - \pi_i^*(g') = c\,(e_{BA}\,l_i(g')) - c\,(e_{BA}\,l_i(g)) - 2e_{BA} + e^* > 0$. □

Proof of Proposition 2. From Proposition 1, it follows immediately that if a given pair of players such that none of the two players is involved in a cooperation commitment form a cooperation commitment, the efficiency of the resulting cooperative network is strictly larger than the efficiency of the current cooperative network. Consequently, for a given n, an efficient cooperative network comprises at least as many cooperation commitments as the unique pairwise stable cooperative network, and both networks are identical when $n = 2$. We now assume that n is an even integer strictly greater than 2 and we consider the pairwise stable cooperative network $g \in G$ such that $l_i(g) = 1$ $\forall\, i \in N$ and $jk \notin g$ where $j, k \in N$ and $j \neq k$. Straightforward computations entail that $\sum_{i \in N} \pi_i^*(g \cup jk) - \sum_{i \in N} \pi_i^*(g) = 2\,ne_{BA} - 2\,(c(2\,e_{BA}) - c(e_{BA}))$. Consequently, if $n > (c(2\,e_{BA}) - c(e_{BA}))\,/e_{BA}$, the efficiency achieved by the cooperative network $g \cup jk$ is strictly greater than the efficiency achieved by the pairwise stable cooperative network g. Obviously, for a given effort cost function, $\exists\, n > 2$ such that the former inequality is true, and, for a given $n > 2$, there exist numerous effort functions such that $n \geq c'(2\,e_{BA})$, i.e., such that the former inequality holds. The same logic can be applied for an odd number of players strictly greater than 2. □

Proof of Proposition 3. Consider the difference $\sum_{i \in N} \pi_i^*(g') - \sum_{i \in N} \pi_i^*(g)$ where g' and g are elements of $G = \{g \mid g \subseteq g^N\}$ such that $g' = g \cup \{ij\}$ $(g \neq g^N)$. As a player's equilibrium payoff in a given cooperative network is a function of the total number of players who have formed at least one cooperation commitment, three cases have to be considered: i) if $l_i(g) \geq 1$ and $l_j(g) \geq 1$ then $n(g') = n(g)$ and $\sum_{i \in N} \pi_i^*(g') - \sum_{i \in N} \pi_i^*(g) = \frac{2}{\mu}\,(n - 2 + 1 - l_i(g) - l_j(g))$; ii) if $l_i(g) = 0$ and $l_j(g) \geq 1$ then $n(g') = n(g) + 1$ and $\sum_{i \in N} \pi_i^*(g') - \sum_{i \in N} \pi_i^*(g) = \frac{1}{2\mu}\left(3(n-2) + \frac{5}{2} - 4l_j(g)\right)$; iii) if $l_i(g) = l_j(g) = 0$ then $n(g') = n(g) + 2$ and $\sum_{i \in N} \pi_i^*(g') - \sum_{i \in N} \pi_i^*(g) = \frac{1}{2\mu}\,(2(n-2) + 1)$.

Two players, such that none of them is currently involved in a cooperation commitment, who form a cooperation commitment with each other increase the efficiency of a cooperative network (case iii). Therefore, the only efficient cooperative network if $n = 2$ is the complete network and at most one player is not involved in any cooperation commitment in an efficient cooperative network. If $n = 3$ then a player who is not involved yet in any cooperation commitment increases the efficiency of the cooperative network by forming a cooperation commitment with a player who is already involved in another cooperation commitment (case ii). If two players who are already involved in a cooperation commitment form a cooperation commitment with each other then the efficiency level of the cooperative network remains constant (case i). Therefore, both the cooperative network with two cooperation commitments and the complete network are efficient in case of 3 players.

We now restrict ourselves to $n \geq 4$. Let $G^l = \{g \mid g \subseteq g^N, l(g) = l, \text{ and } n = n(g)\}$ denote the subset of cooperative networks comprised of $l \geq n-1$ cooperation commitments and in which each player has formed at least one cooperation commitment. If $l = n^2/4$ then the sum of the players' equilibrium effort levels equals ne^{**}, i.e., it corresponds to the efficient total amount of effort. For a given l, if g_1 and g_2 are elements of G^l then $\sum_{i\in N} \pi_i^*(g_1) \geq \sum_{i\in N} \pi_i^*(g_2) \Leftrightarrow \sum_{i\in N} l_i(g_1)^2 \leq \sum_{i\in N} l_i(g_2)^2$. Consequently, if n is an even number then in the unique efficient cooperative network each player forms $n/2$ cooperation commitments which implies that each player exerts an equilibrium effort level of e^{**} (of course, in such a cooperative network the total number of cooperation commitments equals $\frac{n^2}{4}$).

If n is an odd number then it is not possible to achieve full efficiency. According to case i), if g is an efficient cooperative network then, $\forall\, ij \in g$, $n \geq l_i(g) + l_j(g) \geq n-1$ which implies that $\frac{n^2}{4} \geq l(g) \geq \frac{n(n-1)}{4}$. As, for a given l, the most efficient cooperative networks are those which minimize the sum of the squared number of individuals cooperation commitments, $\frac{n+1}{2} \geq l_i(g) \geq \frac{n-1}{2}$ $\forall\, i \in N$ if g is an efficient cooperative network. Among all efficient cooperative networks, the minimally connected cooperative network is made of $\underline{l}$ cooperation commitments where $\underline{l}$ denote the smallest integer in the interval $\left[\frac{n(n-1)}{4}, \frac{n^2}{4}\right]$. □

Proofs of Section 5

General Proposition underlying Proposition 4

Proposition 6 states the set of all possible equilibrium outcomes (including outcomes with extreme preferences which cannot be expected to occur in the laboratory).

Proposition 6 *In the effort game without the possibility to form pairwise commitments:*

- *If* $\alpha_i < \frac{1}{3}$ *and* $\beta_i < \frac{1}{5}$*, then player* i *has a (strictly) dominant strategy, viz.,* $e_i = e^* = 2$.
- *If* $\alpha_1 \geq \frac{1}{3}$ *or* $\beta_1 \geq \frac{1}{5}$*, but* $\alpha_j < \frac{1}{3}$ *and* $\beta_j < \frac{1}{5}$, $\forall j \in \{2,3,4\}$*, then* $(2,2,2,2)$, $\forall i \in \{1,2,3,4\}$ *is the unique equilibrium.*
- *If* $\alpha_i \geq \frac{1}{3}$ *or* $\beta_i \geq \frac{1}{5}$, $\forall i \in \{1,2\}$*, but* $\alpha_j < \frac{1}{3}$ *and* $\beta_j < \frac{1}{5}$, $\forall j \in \{3,4\}$ *then in addition to* $(2,2,2,2)$ *the following are equilibrium outcomes:* ***(a)*** $(3,3,2,2)$ *if* $\beta_i - 2\alpha_i \geq \frac{3}{5}$, $\forall i \in \{1,2\}$, ***(b)*** $(0,0,2,2)$ *if* $\alpha_i - 2\beta_i \geq 9$, $\forall i \in \{1,2\}$, ***(c)*** $(1,1,2,2)$ *if* $\alpha_i - 2\beta_i \geq 1$, $\forall i \in \{1,2\}$.
- *If* $\alpha_i \geq \frac{1}{3}$ *or* $\beta_i \geq \frac{1}{5}$, $\forall i \in \{1,2,3\}$ *but* $\alpha_4 < \frac{1}{3}$ *and* $\beta_4 < \frac{1}{5}$ *then in addition to* $(2,2,2,2)$ *the following are equilibrium outcomes:* ***(a)*** $(3,3,3,2)$ *if* $2\beta_i - \alpha_i \geq \frac{3}{5}$, $\forall i \in \{1,2,3\}$, ***(b)*** $(4,4,4,2)$ *if* $2\beta_i - \alpha_i \geq \frac{9}{7}$, $\forall i \in \{1,2,3\}$, ***(c)*** $(5,5,5,2)$ *if* $2\beta_i - \alpha_i \geq \frac{15}{9}$, $\forall i \in \{1,2,3\}$, ***(d)*** $(6,6,6,2)$ *if* $2\beta_i - \alpha_i \geq \frac{21}{11}$, $\forall i \in \{1,2,3\}$, ***(e)*** $(1,1,1,2)$ *if* $2\alpha_i - \beta_i \geq 1$, $\forall i \in \{1,2,3\}$, ***(f)*** $(0,0,0,2)$ *if* $2\alpha_i - \beta_i \geq 9$, $\forall i \in \{1,2,3\}$, ***(g)*** $(1,0,0,2)$ *if* $1 \leq 2\alpha_1 - \beta_1 \leq 9$ *and* $\alpha_j - 2\beta_j \geq 9$, $\forall j \in \{2,3\}$, ***(h)*** $(1,1,2,2)$ *if* $\alpha_i - 2\beta_i \geq 1$, $\forall i \in \{1,2\}$ *and* $2\alpha_3 - \beta_3 \leq 1$, ***(i)*** $(0,0,2,2)$ *if* $\alpha_i - 2\beta_i \geq 9$, $\forall i \in \{1,2\}$ *and* $2\alpha_3 - \beta_3 \leq 3$.
- *If* $\alpha_i \geq \frac{1}{3}$ *or* $\beta_i \geq \frac{1}{5}$, $\forall i \in \{1,2,3,4\}$*, then in addition to* $(2,2,2,2)$ *the following are equilibrium outcomes:* ***(a)*** (e_i, e_i, e_i, e_i) *with* $e_i \geq 3$ *if* $\beta_i \geq \frac{2e_i-5}{2e_i-1}$, $\forall i \in \{1,2,3,4\}$ ***(b)*** $(e_i, e_i, e_i, 2)$ *with*

*$e_i \geq 3$ if $\alpha_i - 2\beta_i \leq \frac{15-6e_i}{2e_i-1}$, $\forall i \in \{1,2,3\}$, and if $\beta_4 \leq \frac{1}{3}$, **(c)** $(1,1,1,1)$ if $\alpha_i \geq 1$, $\forall i \in \{1,2,3,4\}$, **(d)** $(0,0,0,0)$ if $\alpha_i \geq 9$, $\forall i \in \{1,2,3,4\}$, **(e)** $(1,0,0,0)$ if $\frac{1}{3} \leq \alpha_1 \leq 3$ and $2\alpha_j - \beta_j \geq 9$, $\forall i \in \{2,3,4\}$, **(f)** $(1,1,0,0)$ if $2\alpha_i - \beta_i \leq 9$ and $\alpha_i \geq \frac{1}{3}$, $\forall i \in \{1,2\}$ and $\alpha_j - 2\beta_j \geq 9$, $\forall j \in \{3,4\}$, **(g)** $(2,1,1,1)$ if $\alpha_1 \leq \frac{1}{3}$ and $2\alpha_j - \beta_j \geq 1$, $\forall j \in \{2,3,4\}$, **(h)** $(2,0,0,0)$ if $\alpha_1 \leq 1$ and $2\alpha_j - \beta_j \geq 9$, $\forall j \in \{2,3,4\}$, **(i)** $(2,1,0,0)$ if $\alpha_1 \leq \frac{3}{31}$ and $1 \leq 2\alpha_2 - \beta_2 \leq 9$ and $4\alpha_j - 2\beta_j \geq 9$, $\forall j \in \{3,4\}$, **(j)** $(2,2,1,1)$ if $2\alpha_i - \beta_i \leq 1$, $\forall i \in \{1,2\}$ and $\alpha_j - 2\beta_j \geq 1$, $\forall j \in \{3,4\}$, and finally **(k)** $(2,2,0,0)$ if $2\alpha_i - \beta_i \leq 3$, $\forall i \in \{1,2\}$ and $\alpha_j - 2\beta_j \geq 9$, $\forall j \in \{3,4\}$.*

Proof

Dominant strategy

CASE 1: Let assume player i is player 1 and that $e_1 = 2$. Assume player 1 is comparatively disadvantaged. The inequity experienced by player 1 is maximized in the case in which $e_j = 0$, $\forall j \in \{2,3,4\}$. Player 1 deviates to $e_1 = 1$ only if $\alpha_1 < \frac{1}{3}$.

CASE 2: Let assume $e_1 = 2$ and player 1 is comparatively advantaged. The inequity experienced by player 1 is maximized in the case in which $e_j = 12$, $\forall j \in \{2,3,4\}$. Player 1 increases his effort to $e_1 = 3$ if $\beta_1 < \frac{1}{5}$.

Two fair players

The following cases can be distinguished (where $e_3^* = e_4^* = 2$):

Case 1: $e_1 = e_2 > 2$
Case 2: $2 > e_1 = e_2$
Case 3: $e_1 > e_2 > 2$
Case 4: $2 > e_1 > e_2$
Case 5: $e_1 > 2 > e_2$
Case 6: $e_1 > 2 = e_2$
Case 7: $e_1 = 2 > e_2$
Case 8: $e_1 = e_2 = 2$

CASE 8 is always an equilibrium. Some of the other cases can trivially be ruled out as potential equilibria:

CASE 3, 5, 6: Player 1 has an incentive to decrease his effort as this leads to an increase in his material payoff while it reduces his inequity w.r.t. all other players.

CASE 4, 7: Player 2 has an incentive to increase his effort as this leads to an increase in his material payoff while it reduces his inequity w.r.t. all other players.

CASE 1: Let focus on player 1 (where the argumentation applies also to player 2). We will in the following denote the status quo effort of the player under consideration by e_0 and his final effort after deviation by e_f. The following deviations are possible: A) increase, where $e_f > e_0$, B) decrease, where $e_f \geq 2$, and C) decrease, where $e_f < 2$. A) and C) have been ruled out before. In B) the following subcases can be distinguished:

- $e_1 = e_0 \geq 4$ and $e_f \geq 3$: A deviation from e_0 to e_f is not profitable for player 1 if $U_1(e_0, e_0, 2, 2) \geq U_1(e_f, e_0, 2, 2)$, viz., if $\frac{12-3(e_0+e_f)}{e_0+e_f} \geq 2\alpha_1 - \beta_1$. This condition cannot be fulfilled.

- $e_1 = e_0 \geq 3$ and $e_f = 2$: The case in which $e_0 \geq 4$ can be excluded as we have shown before that a deviation from $e_0 \geq 4$ to $e_f = 3$ is always profitable. A deviation from $e_0 = 3$ to $e_f = 2$ is not profitable for player 1 if $U_1(3, 3, 2, 2) \geq U_1(2, 3, 2, 2)$, viz., if $\beta_1 - 2\alpha_1 \geq \frac{3}{5}$.

CASE 2: Let focus on player 1 (where the argumentation applies also to player 2). The following deviations are possible: A) increase, where $e_f \leq 2$, B) increase, where $e_f > 2$, and C) decrease, where $e_f = 0$ if $e_1 = 1$. B) and C) have been ruled out before.

- $e_1 = e_0 = 0$ and $e_f = 1$: A deviation from e_0 to e_f is not profitable for player 1 if $U_1(0,0,2,2) \geq U_1(1,0,2,2)$, viz., if $\alpha_1 - 2\beta_1 \geq 9$.

- $e_1 = e_0 = 0$ and $e_f = 2$: A deviation from e_0 to e_f is not profitable for player 1 if $U_1(0,0,2,2) \geq U_1(2,0,2,2)$, viz., if $\alpha_1 - 2\beta_1 \geq 3$. This condition is subsumed by the previous condition.

- $e_1 = e_0 = 1$ and $e_f = 2$: A deviation from e_0 to e_f is not profitable for player 1 if $U_1(1,1,2,2) \geq U_1(2,1,2,2)$, viz., $\alpha_1 - 2\beta_1 \geq 1$.

Three fair players

The following cases can be distinguished (where $e_4^* = 2$):

Case 1: $e_1 = e_2 = e_3 > 2$
Case 2: $e_1 = e_2 > e_3 > 2$
Case 3: $e_1 > e_2 = e_3 > 2$
Case 4: $e_1 > e_2 > e_3 > 2$
Case 5: $e_1 = e_2 > e_3 = 2$
Case 6: $e_1 > e_2 > e_3 = 2$
Case 7: $e_1 > e_2 = e_3 = 2$
Case 8: $e_1 = e_2 = e_3 < 2$
Case 9: $2 > e_1 = 1 > e_2 = e_3 = 0$
Case 10: $2 > e_1 = e_2 = 1 > e_3 = 0$
Case 11: $e_1 = e_2 < e_3 = 2$
Case 12: $2 = e_1 > e_2 = 1 > e_3 = 0$
Case 13: $2 = e_1 = e_2 > e_3$
Case 14: $e_1 = e_2 = e_3 = 2$
Case 15: $e_1 = e_2 > 2 > e_3$
Case 16: $e_1 > e_2 > 2 > e_3$
Case 17: $e_1 > e_2 = 2 > e_3$
Case 18: $e_1 > 2 > e_2 = e_3$
Case 19: $e_1 > 2 > e_2 = 1 > e_3 = 0$

CASE 14 is an equilibrium. The following cases can be straightforwardly excluded as potential equilibria:
CASE 2: Player 1 has an incentive to decrease his effort (given $\alpha_1 \geq \beta_1$) as this increases his material payoff and decreases his relative disadvantage w.r.t. 3 other players while it only increases his relative advantage w.r.t. one player.
CASE 3, 4, 6, 7, 17, 18, 19, 20: Player 1 has an incentive to decrease his effort as this leads to an increase in his material payoff while it reduces his inequity w.r.t. all other players.
CASE 10, 12, 13, 16, 17, 18, 20: Player 3 has an incentive to increase his effort as this leads to an increase in his material payoff while it reduces his inequity w.r.t. all other players.

CASE 1: Let focus on player 1 (where the argumentation applies also to players 2 and 3). The following deviations are possible A) increase, where $e_f \geq e_0$, B) decrease, where $e_0 > e_f \geq 2$, and C) decrease, where $e_f < 2 < e_0$. A) and C) have been ruled out before. In B) the following subcases can be distinguished:

- $e_1 = e_0 \geq 3$ and $e_f = 2$: A deviation from e_0 to e_f is not profitable for player 1 if $U_1(e_0, e_0, e_0, 2) \geq U_1(2, e_0, e_0, 2)$, viz., if $2\beta_1 - \alpha_1 \geq \frac{3}{5}$.

- $e_1 = e_0 \geq 4$ and $e_f \geq 3$: A deviation from e_0 to e_f is not profitable for player 1 if $U_1(e_0, e_0, e_0, 2) \geq U_1(e_f, e_0, e_0, 2)$, viz., if $2\beta_1 - \alpha_1 \geq \frac{3(e_f+e_0)-12}{e_0+e_f}$.

CASE 5: Let focus on player 1 (where the argumentation applies identically to player 2): Player 1 could A) increase, where $e_f \geq e_0$, B) decrease, where $e_0 > e_f \geq 2$, and C) decrease, where $e_f < 2 < e_0$. A) and C) have been ruled out before.

- $e_1 = e_0 \geq 4$ and $e_f \geq 3$: A deviation from e_0 to e_f is not profitable for player 1 if $U_1(e_0, e_0, 2, 2) \geq U_1(e_f, e_0, 2, 2)$, viz., if $2\alpha_1 - \beta_1 \leq \frac{12}{e_0+e_f} - 3$. As this condition cannot be fulfilled, $(e_0, e_0, 2, 2)$ is no equilibrium if $e_0 \geq 4$.

- $e_1 = e_2 = e_0 \geq 3$ and $e_f = 2$: A deviation from e_0 to e_f is not profitable for player 1 if $U_1(e_0, e_0, 2, 2) \geq U_1(2, e_0, 2, 2)$, viz., if $2\beta_1 - \alpha_1 \leq \frac{12}{e_0+2} - 3$. For $e_0 \geq 3$ this condition cannot be fulfilled. Hence, $(e_0, e_0, 2, 2)$ is also no equilibrium if $e_0 = 3$.

$\rightarrow$ Case 5 cannot be an equilibrium as player i, $\forall i \in \{1, 2\}$ always as an incentive to decrease his effort.

CASE 8: Let focus on player 1 (where the argumentation applies identically to players 2 and 3): Player 1 could A) increase, where $e_0 < e_f \leq 2$, B) increase, where $e_0 < 2 < e_f$, and C) decrease, where $e_f < e_0 < 2$. B) and C) have been ruled out before.

- $e_1 = e_0 = 1$ and $e_f = 2$: A deviation from e_0 to e_f is not profitable for player 1 if $U_1(1, 1, 1, 2) \geq U_1(2, 1, 1, 2)$, viz., if $2\alpha_1 - \beta_1 \geq 1$.

- $e_1 = e_0 = 0$ and $e_f = 1$: A deviation from e_0 to e_f is not profitable for player 1 if $U_1(0, 0, 0, 2) \geq U_1(1, 0, 0, 2)$, viz., if $2\alpha_1 - \beta_1 \geq 9$. This condition is binding (it subsumes the condition for $e_f = 2$).

CASE 9:

SUBCASE 9A) Let focus on player 2 (where the argumentation applies identically to player 3): Player 2 could A) increase, where $e_0 = 0 < e_f \leq 2$, B) increase, where $e_0 < 2 < e_f$. Case B) has been ruled out before.

- $e_2 = e_0 = 0$ and $e_f = 1$: A deviation from e_0 to e_f is not profitable for player 2 if $U_2(1, 0, 0, 2) \geq U_2(1, 1, 0, 2)$, viz., if $\alpha_2 - 2\beta_2 \geq 9$. This condition is binding (it subsumes the condition for $e_f = 2$).

SUBCASE 9B) Let focus on player 1. Player 1 could A) increase, where $e_0 = 1 < e_f = 2$, B) increase, where $e_0 = 1 < 2 < e_f$, and C) decrease, where $e_0 = 1 > e_f = 0$. Case B) has been ruled out before.

- $e_1 = e_0 = 1$ and $e_f = 2$: A deviation from e_0 to e_f is not profitable for player 1 if $U_1(1, 0, 0, 2) \geq U_1(2, 0, 0, 2)$, viz., if $2\alpha_1 - \beta_1 \geq 1$.

- $e_1 = e_0 = 1$ and $e_f = 0$: A deviation from e_0 to e_f is not profitable for player 1 if $U_1(1, 0, 0, 2) \geq U_1(0, 0, 0, 2)$, viz., if $2\alpha_1 - \beta_1 \leq 9$.

$\rightarrow$ Player 1 does not deviate from $(1, 0, 0, 2)$, if $1 \leq 2\alpha_1 - \beta_1 \leq 9$.

CASE 11:

SUBCASE 11A) Let focus on player 1 (where the argumentation applies also to player 2): Player

1 could A) increase, where $e_0 < e_f \leq 2$, B) increase, where $e_0 < 2 < e_f$, and C) decrease, where $e_0 > e_f$. Cases B) and C) have been ruled out before.

- $e_1 = e_0 \leq 1$ and $e_f = 2$: A deviation from e_0 to e_f is not profitable for player 1 if $U_1(e_0, e_0, 2, 2) \geq U_1(2, e_0, 2, 2))$, viz., if $\frac{6-3e_0}{e_0+2} \leq \alpha_1 - 2\beta_1$. For $e_1 = 1$ it follows that $\alpha_1 - 2\beta_1 \geq 1$ and for $e_1 = 0$ it follows that $\alpha_1 - 2\beta_1 \geq 3$.

- $e_1 = e_0 = 0$ and $e_f = 1$: A deviation from e_0 to e_f is not profitable for player 1 if $U_1(0, 0, 2, 2) \geq U_1(1, 0, 2, 2)$, viz., if $\alpha_1 - 2\beta_1 \geq 9$.

Subcase 11b) Let focus on player 3. Player 3 could A) increase, where $e_0 = 2 < e_f$, B) decrease, where $e_0 = 2 > e_f \geq e_1 = e_2$, and C) decrease, where $e_0 = 2 > e_1 = e_2 > e_f$. Cases A) and C) have been ruled out before.

- $e_3 = 2 = e_0$ and $e_f = e_1 = e_2$: A deviation from e_0 to e_f is not profitable for player 3 if $U_3(e_f, e_f, 2, 2) \geq U_3(e_f, e_f, e_f, 2)$, viz., if $\frac{6-3e_f}{2+e_f} \geq 2\alpha_3 - \beta_3$. For $e_f = 0$ this leads to $2\alpha_3 - \beta_3 \leq 3$, and for $e_f = 1$ this leads to $2\alpha_3 - \beta_3 \leq 1$.

- $e_3 = 2 = e_0$ and $e_f = 1 > e_1 = e_2 = 0$: A deviation from e_0 to e_f is not profitable for player 3 if $U_3(0, 0, 2, 2) \geq U_3(0, 0, 1, 2)$, viz., if $2\alpha_3 - \beta_3 \leq 1$.

Four fair players

The following cases can be distinguished:

Case 1: $e_1 = e_2 = e_3 = e_4 > 2$
Case 2: $e_1 = e_2 = e_3 > e_4 > 2$
Case 3: $e_1 > e_2 = e_3 = e_4 > 2$
Case 4: $e_1 = e_2 > e_3 = e_4 > 2$
Case 5: $e_1 = e_2 > e_3 > e_4 > 2$
Case 6: $e_1 > e_2 = e_3 > e_4 > 2$
Case 7: $e_1 > e_2 > e_3 = e_4 > 2$
Case 8: $e_1 > e_2 > e_3 > e_4 > 2$
Case 9: $e_1 = e_2 = e_3 > e_4 = 2$
Case 10: $e_1 = e_2 > e_3 > e_4 = 2$
Case 11: $e_1 > e_2 = e_3 > e_4 = 2$
Case 12: $e_1 > e_2 > e_3 > e_4 = 2$
Case 13: $e_1 = e_2 > e_3 = e_4 = 2$
Case 14: $e_1 > e_2 > e_3 = e_4 = 2$
Case 15: $e_1 > e_2 = e_3 = e_4 = 2$
Case 16: $e_1 = e_2 = e_3 = e_4 = 2$
Case 17: $e_1 = e_2 = e_3 = e_4 < 2$
Case 18: $2 > e_1 = e_2 = e_3 = 1 > e_4 = 0$
Case 19: $2 > e_1 = 1 > e_2 = e_3 = e_4 = 0$
Case 20: $2 > e_1 = e_2 = 1 > e_3 = e_4 = 0$
Case 21: $2 = e_1 > e_2 = e_3 = e_4$
Case 22: $2 = e_1 > e_2 = e_3 = 1 > e_4 = 0$
Case 23: $2 = e_1 > e_2 = 1 > e_3 = e_4 = 0$
Case 24: $2 = e_1 = e_2 > e_3 = e_4$
Case 25: $2 = e_1 = e_2 > e_3 = 1 > e_4 = 0$
Case 26: $2 = e_1 = e_2 = e_3 > e_4$
Case 27: $e_1 = e_2 = e_3 > 2 > e_4$
Case 28: $e_1 = e_2 > e_3 > 2 > e_4$
Case 29: $e_1 > e_2 = e_3 > 2 > e_4$
Case 30: $e_1 > e_2 > e_3 > 2 > e_4$
Case 31: $e_1 = e_2 > e_3 = 2 > e_4$
Case 32: $e_1 > e_2 > e_3 = 2 > e_4$
Case 33: $e_1 = e_2 > 2 > e_3 = e_4$
Case 34: $e_1 > e_2 > 2 > e_3 = e_4$
Case 35: $e_1 = e_2 > 2 > e_3 = 1 > e_4 = 0$
Case 36: $e_1 > e_2 > 2 > e_3 > e_4$
Case 37: $e_1 > e_2 = e_3 = 2 > e_4$
Case 38: $e_1 > e_2 = 2 > e_3 = e_4$
Case 39: $e_1 > e_2 = 2 > e_3 = 1 > e_4 = 0$
Case 40: $e_1 > 2 > e_2 = e_3 = e_4$
Case 41: $e_1 > 2 > e_2 = e_3 = 1 > e_4 = 0$
Case 42: $e_1 > 2 > e_2 = 1 > e_3 = e_4 = 0$

CASE 16 is an equilibrium. The following cases can be straightforwardly excluded as potential equilibria:

CASE 3, 6-8, 11-12, 14, 15, 29, 30, 32, 34, AND 36-42: Player 1 has an incentive to deviate by decreasing his effort as this increases his material payoff and decreases inequity w.r.t. all other players.

CASE 18, 22, 25-32, 35-37, 39, AND 41: Player 4 has an incentive to deviate by increasing his effort as this increases his material payoff and decreases inequity w.r.t. all other players.

CASE 4, 5, 10, 13, 33: Player 1 has an incentive to decrease his effort (given $\alpha_1 \geq \beta_1$) as this increases his material payoff and decreases his relative disadvantage w.r.t. 3 other players while it only increases his relative advantage w.r.t. one player.

CASE 1: Let assume that player i is player 1. The following deviations are possible A) increase, where $e_f > e_0 > 2$, B) decrease, where $e_0 > e_f \geq 2$, and C) decrease, where $e_f < 2$. A) and C) have been ruled out before.

- $e_1 = e_0 \geq 3$ and $e_f \geq 2$: A deviation from e_0 to e_f is not profitable for player 1 if $U_1(e_0, e_0, e_0, e_0) \geq U_1(e_f, e_0, e_0, e_0)$, viz., if $\beta_1 \geq \frac{(e_0+e_f)-4}{e_0+e_f}$. The condition is binding for $e_f = e_0 - 1$.

CASE 2: Let focus on player 4. The following deviations are possible A) increase, where $e_1 = e_2 = e_3 \geq e_f > e_0$, B) increase, where $e_f > e_1 = e_2 = e_3 > e_0$, C) decrease, where $e_0 > e_f \geq 2$, and D) decrease, where $e_f < 2 < e_0$. B) and D) have been ruled out before.

- $e_1 = e_2 = e_3 > e_f > e_0 = e_4 > 2$: A deviation from e_0 to e_f is not profitable for player 4 if $U_4(e_1, e_2, e_3, e_0) \geq U_4(e_1, e_2, e_3, e_f)$, viz., if $\beta_4 \leq \frac{e_0+e_f-4}{e_0+e_f}$. The condition is binding for $e_f = e_0 + 1$ and subsumes the case $e_f = e_1 = e_2$.

- $e_4 = e_0 > e_f \geq 2$: A deviation from e_0 to e_f is not profitable for player 4 if $U_4(e_1, e_2, e_3, e_0) \geq U_4(e_1, e_2, e_3, e_f)$, viz., if $\beta_4 \geq \frac{e_0+e_f-4}{e_0+e_f}$. The condition is binding for $e_f = e_0 - 1$.

$\rightarrow$ As the two conditions cannot be fulfilled simultaneously, player 4 always deviates. Consequently, case 2 cannot be an equilibrium.

CASE 9:

SUBCASE 9A) Let focus on player 4: The following deviations are possible A) increase, where $e_1 = e_2 = e_3 \geq e_f > e_0 = 2$, B) increase, where $e_f > e_1 = e_2 = e_3 > e_0 = 2$, C) decrease, where $e_0 = 2 > e_f$. B) and C) have been ruled out before.

- $e_1 = e_2 = e_3 \geq e_f > e_0 = e_4 = 2$: A deviation from e_0 to e_f is not profitable for player 4 if $U_4(e_1, e_2, e_3, 2) \geq U_4(e_1, e_2, e_3, e_f)$, viz., if $\beta_4 \leq \frac{e_f-2}{e_f}$. The condition is binding for $e_f = e_0 + 1 = 3$, i.e., $\beta_4 \leq \frac{1}{3}$.

SUBCASE 9B) Let focus on player 1 (where the same argumentation applies to players 2, and 3): The following deviations are possible A) increase, where $e_f > e_0 = e_1$, B) decrease, where $e_0 > e_f \geq e_4 = 2$, and C) decrease, where $e_f < 2$. A) and C) have been ruled out before.

- $e_0 > e_f \geq e_4 = 2$: A deviation from e_0 to e_f is not profitable for player 1 if $U_1(e_1, e_2, e_3, 2) \geq U_1(e_f, e_2, e_3, 2)$, viz., if $\alpha_1 - 2\beta_1 \leq \frac{12-3(e_0+e_f)}{e_0+e_f}$. The condition is binding for $e_f = e_0 - 1$.

Case 17: Let focus on player 1 (the argumentation applies to all players). The following deviations are possible A) increase, where $e_0 < e_f \leq 2$, B) increase, where $e_f > 2$, C) decrease, where $2 > e_0 > e_f$. B) and C) have been ruled out before.

- $e_1 = e_0 < e_f \leq 2$. A deviation from e_0 to e_f is not profitable for player 4 if $U_4(e_0, e_0, e_0, e_0) \geq U_4(e_0, e_0, e_0, e_f)$, viz., if $\alpha_1 \geq \frac{12-3(e_0+e_f)}{e_0+e_f}$. The condition is binding for $e_f = e_0 + 1$. For $e_0 = 1$ this leads to $\alpha_1 \geq 1$ and for $e_0 = 0$ this leads to $\alpha_1 \geq 9$.

Case 19:

Subcase 19a) Let focus on player 1. The following deviations are possible A) increase, where $e_f = 2$, B) increase, where $e_f > 2$, C) decrease, where $e_f = 0$. B) has been ruled out before.

- $e_1 = e_0 = 1$ and $e_f = 0$: A deviation from e_0 to e_f is not profitable for player 1 if $U_1(1,0,0,0) \geq U_1(0,0,0,0)$, viz., if $\alpha_1 \leq 3$.

- $e_1 = e_0 = 1$ and $e_f = 2$: A deviation from e_0 to e_f is not profitable for player 1 if $U_1(1,0,0,0) \geq U_1(2,0,0,0)$, viz., if $\alpha_1 \geq \frac{1}{3}$.

Subcase 19b) Let focus on player 2 (where the same argumentation applies to players 2, 3 and 4): The following deviations are possible A) increase, where $e_f \leq 2$, and B) increase, where $e_f > 2$. B) has been ruled out before.

- $e_0 = 0$ and $e_f = 1$ (binding condition): A deviation from e_0 to e_f is not profitable for player 2 if $U_2(1,0,0,0) \geq U_2(1,1,0,0)$, viz., if $2\alpha_2 - \beta_2 \geq 9$.

Case 20:

Subcase 20a) Let focus on player 1 (where the same argumentation applies to player 2): The following deviations are possible A) increase, where $e_f = 2$, B) increase, where $e_f > 2$, C) decrease, where $e_f = 0$. B) has been ruled out before.

- $e_1 = e_0 = 1$ and $e_f = 0$: A deviation from e_0 to e_f is not profitable for player 1 if $U_1(1,1,0,0) \geq U_1(0,1,0,0)$, viz., if $2\alpha_1 - \beta_1 \leq 9$.

- $e_1 = e_0 = 1$ and $e_f = 2$: A deviation from e_0 to e_f is not profitable for player 1 if $U_1(1,1,0,0) \geq U_1(2,1,0,0)$, viz., if $\alpha_1 \geq \frac{1}{3}$.

Subcase 20b) Let focus on player 3 (where the same argumentation applies to player 4): The following deviations are possible A) increase, where $e_f \leq 2$, and B) increase, where $e_f > 2$. B) has been ruled out before.

- $e_3 = e_0 = 0$ and $e_f = 1$ (binding condition): A deviation from e_0 to e_f is not profitable for player 3 if $U_3(1,1,0,0) \geq U_3(1,1,1,0)$, viz., if $\alpha_3 - 2\beta_3 \geq 9$.

Case 21:

Subcase 21a) Let focus on player 1: The following deviations are possible A) increase, where $e_f > 2$, C) decrease, where $e_f \geq e_2 = e_3 = e_4$, and C) decrease, where $e_f = 0 < e_2 = e_3 = e_4$. A) and C) have been ruled out before.

- $e_1 = e_0 = 2$ and $e_f = 1 > e_2 = e_3 = e_4 = 0$: A deviation from e_0 to e_f is not profitable for player 1 if $U_1(2,0,0,0) \geq U_1(1,0,0,0)$, viz., if $\alpha_1 \leq \frac{1}{3}$.

- $e_1 = e_0 = 2$ and $e_f = e_2 = e_3 = e_4$: A deviation from e_0 to e_f is not profitable for player 1 if $U_1(2, e_f, e_f, e_f) \geq U_1(e_f, e_f, e_f, e_f)$, viz., if $\alpha_1 \leq \frac{2-e_f}{2+e_f}$. With $e_f = 1$, $\alpha_1 \leq \frac{1}{3}$ and with $e_f = 0$,

$\alpha_1 \leq 1$.

Subcase 21b) Let focus on player 2 (where the same argumentation applies to players 3 and 4): The following deviations are possible A) increase, where $e_f \leq 2$, and B) increase, where $e_f > 2$, and C) decrease. B) and C) have been ruled out before.

- e_0 and $e_f = e_0 + 1$ (binding condition): A deviation from e_0 to e_f is not profitable for player 2 if $U_2(2, e_0, e_0, e_0) \geq U_2(2, e_f, e_0, e_0)$, viz., if $2\alpha_2 - \beta_2 \geq \frac{12-3(e_f+e_0)}{e_f+e_0}$. With $e_0 = 1$, $2\alpha_2 - \beta_2 \geq 1$, and with $e_0 = 0$, $2\alpha_2 - \beta_2 \geq 9$.

Case 23:

Subcase 23a) Let focus on player 1: The following deviations are possible A) increase, where $e_f > 2$, and B) decrease, where $e_f \leq 1$. A) has been ruled out before.

- $e_1 = e_0 = 2$ and $e_f = 1$: A deviation from e_0 to e_f is not profitable for player 1 if $U_1(2, 1, 0, 0) \geq U_1(1, 1, 0, 0)$, viz., if $\alpha_1 \leq \frac{3}{31}$.

- $e_1 = e_0 = 2$ and $e_f = 0$: A deviation from e_0 to e_f is not profitable for player 1 if $U_1(2, 1, 0, 0) \geq U_1(0, 1, 0, 0)$, viz., if $33\alpha_1 - \beta_1 \leq 12$. $\rightarrow$ Player 1 does not deviate from $(2, 1, 0, 0)$ if $\alpha_1 \leq \frac{3}{31}$ as the second condition is trivially fulfilled as $\alpha_1 \leq \frac{3}{31} \Rightarrow \beta_1 \geq \frac{1}{5}$.

Subcase 23b) Let focus on player 2: The following deviations are possible A) increase, where $e_f = 2$, B) increase, where $e_f > 2$, C) decrease, where $e_f = 0$. B) has been ruled out before.

- $e_0 = 1$ and $e_f = 2$: A deviation from e_0 to e_f is not profitable for player 2 if $U_2(2, 1, 0, 0) \geq U_2(2, 2, 0, 0)$, viz., if $2\alpha_2 - \beta_2 \geq 1$.

- $e_0 = 1$ and $e_f = 0$: A deviation from e_0 to e_f is not profitable for player 2 if $U_2(2, 1, 0, 0) \geq U_1(2, 0, 0, 0)$, viz., if $2\alpha_2 - \beta_2 \leq 9$.

Subcase 23c) Let focus on player 3 (where the same argumentation applies to player 4): The following deviations are possible A) increase, where $e_f \leq 2$, and B) increase, where $e_f > 2$. B) has been ruled out before.

- $e_0 = 0$ and $e_f = 1$ (binding condition): A deviation from e_0 to e_f is not profitable for player 3 if $U_3(2, 1, 0, 0) \geq U_3(2, 1, 1, 0)$, viz., if $4\alpha_3 - 2\beta_3 \geq 9$.

Case 24:

Subcase 24a) Let focus on player 1 (where the same argumentation applies to player 2): The following deviations are possible A) increase, where $e_f > 2$, B) decrease, where $e_f \geq e_3 = e_4$, C) decrease, where $e_f < e_3 = e_4$. A) and C) have been ruled out before.

- $e_0 = 2$ and $e_f = e_3 = e_4$: A deviation from e_0 to e_f is not profitable for player 1 if $U_1(2, 2, e_f, e_f) \geq U_1(e_f, 2, e_f, e_f)$, viz., if $2\alpha_1 - \beta_1 \leq \frac{6-3e_f}{2+e_f}$. With $e_f = 1$, $2\alpha_1 - \beta_1 \leq 1$ and with $e_f = 1$, $2\alpha_1 - \beta_1 \leq 3$.

- $e_0 = 2$ and $e_f = 1$ and $e_3 = e_4 = 0$: A deviation from e_0 to e_f is not profitable for player 1 if $U_1(2, 2, 0, 0) \geq U_1(1, 2, 0, 0)$, viz., if $2\alpha_1 - \beta_1 \leq 1$.

Subcase 24b) Let focus on player 3 (where the same argumentation applies to player 4): The following deviations are possible A) increase, where $e_f \leq 2$, B) increase, where $e_f > 2$, C) decrease, where $e_f = 0$. B) and C) have been ruled out before.

- $e_3 = e_0 = 1$ and $e_f = 2$: A deviation from e_0 to e_f is not profitable for player 3 if $U_3(2, 2, 1, 1) \geq U_3(2, 2, 2, 1)$, viz., if $\alpha_3 - 2\beta_3 \geq 1$.

- $e_3 = e_0 = 0$ and $e_f = 1$ (binding case): A deviation from e_0 to e_f is not profitable for player 3 if $U_3(2,2,0,0) \geq U_3(2,2,1,0)$, viz., if $\alpha_3 - 2\beta_3 \geq 9$. □

Proof of Proposition 5

Network-contingent dominant strategies

(i) If $l_i(g) = 0$ and $g = g^{\emptyset}$, then player i has a strictly dominant strategy if $\alpha_i < \frac{1}{3}$ and $\beta_i < \frac{1}{5}$, viz., $e_i^*(g) = 2$.

(i) If $l_i(g) = 0$ and $g = g^1$, then player i has a strictly dominant strategy if $\alpha_i < \frac{7}{5}$ and $\beta_i < \frac{1}{5}$, viz., $e_i^*(g) = 2$.

(ii) If $l_i(g) = 0$ and $g \neq g^1$, then player i has a strictly dominant strategy if $\beta_i < \frac{1}{5}$, viz., $e_i^*(g) = 2$.

(iii) If $l_i(g) = 1$, then player i has a strictly dominant strategy if $\beta_i < \frac{5}{9}$, viz., $e_i^*(g) = 4$.

(iv) If $l_i(g) = 2$, then player i has a strictly dominant strategy if $\beta_i < \frac{13}{17}$, viz., $e_i^*(g) = 8$.

Network-contingent equilibrium efforts with one fair player

Assume player 1 is the fair player. It is trivial to show that the fair player does not move beyond the committed level in symmetric networks.

Case 1: In any network in which player 1 has formed fewer commitments than any other player, and where either $l_3(g) > l_2(g) = l_4(g) > l_1(g)$ or $l_2(g) = l_3(g) = l_4(g) > l_1(g)$, in equilibrium it has to hold that $\frac{5-2e_1}{1-2e_1} \leq \beta_1 \leq \frac{2e_1-3}{2e_1+1}$, as otherwise player 1 would have an incentive to increase/decrease his effort to $e_1 + 1/e_1 - 1$.

Case 2a: In any network in which player 1 and one of the other players have formed an identical number of commitments, and where $l_2(g) = l_3(g) > l_4(g) = l_1(g)$, in equilibrium it has to hold that $\frac{15-6e_1}{1-2e_1} \leq 2\beta_1 - \alpha_1 \leq \frac{6e_1-9}{2e_1+1}$, as otherwise player 1 would have an incentive to increase/decrease his effort to $e_1 + 1/e_1 - 1$.

Case 2b: In any network in which player 1 and one of the other players have formed an identical number of commitments, and where $l_2(g) > l_3(g) = l_1(g) > l_4(g)$, in equilibrium it has to hold that $\alpha_1 - \beta_1 \geq \frac{3(3-2e_1)}{2e_1+1}$ and that $2\alpha_1 - \beta_1 \leq \frac{6e_1-15}{1-2e_1}$. As player 1 is connected, $e_1 \geq 4$. This implies that the first conditions is always and the second never fulfilled. In consequence, player 1 sticks to the level he committed to.

Case 3: In any network in which player 1 and two of the other players have formed an identical number of commitments, and where $l_3(g) > l_2(g) = l_4(g) = l_1(g)$ in equilibrium it has to hold that $\beta_1 \leq \frac{3(2e_1-3)}{2e_1+1}$ and that $2\alpha_1 - \beta_1 \leq \frac{6e_1-15}{1-2e_1}$. As player 1 is connected, $e_1 \geq 4$. This implies that the first conditions is always and the second never fulfilled. In consequence, player 1 sticks to the level he committed to.

We have summarized all resulting network-contingent equilibria in Table 20 (note that some of them are not obtainable with realistic preferences).[40]

[40] All possible outcomes are generated by relabelling players 2, 3, and 4. Equilibria are stated as $(e_1^*(g), e_2^*(g), e_3^*(g), e_4^*(g))$. Due to space constraints $l_i(g)$ is abbreviated as l_i.

Case	Network	Position of fair player	Equilibrium	Condition
I. $l_3 > l_2 = l_4 > l_1$	$g^{2\star}$	$l_1 = 0$	(2, 4, 8, 4)	$\beta_1 \leq \frac{1}{5}$
			(3, 4, 8, 4)	$\frac{1}{5} \leq \beta_1 \leq \frac{3}{7}$
			(4, 4, 8, 4)	$\beta_1 \geq \frac{3}{7}$
	$g^{\star+1}$	$l_1 = 1$	(4, 8, 12, 8)	$\beta_1 \leq \frac{5}{9}$
			(5, 8, 12, 8)	$\frac{5}{9} \leq \beta_1 \leq \frac{7}{11}$
			(6, 8, 12, 8)	$\frac{7}{11} \leq \beta_1 \leq \frac{9}{13}$
			(7, 8, 12, 8)	$\frac{9}{13} \leq \beta_1 \leq \frac{11}{15}$
			(8, 8, 12, 8)	$\beta_1 \geq \frac{11}{15}$
II. $l_2 = l_3 = l_4 > l_1$	$g^{\triangle}$	$l_1 = 0$	(2, 8, 8, 8)	$\beta_1 \leq \frac{1}{5}$
			(3, 8, 8, 8)	$\frac{1}{5} \leq \beta_1 \leq \frac{3}{7}$
			(4, 8, 8, 8)	$\frac{3}{7} \leq \beta_1 \leq \frac{5}{9}$
			(5, 8, 8, 8)	$\frac{5}{9} \leq \beta_1 \leq \frac{7}{11}$
			(6, 8, 8, 8)	$\frac{7}{11} \leq \beta_1 \leq \frac{9}{13}$
			(7, 8, 8, 8)	$\frac{9}{13} \leq \beta_1 \leq \frac{11}{15}$
			(8, 8, 8, 8)	$\beta_1 \geq \frac{11}{15}$
III. $l_2 = l_3 > l_1 = l_4$	g^1	$l_1 = 0$	(2, 4, 4, 2)	$2\beta_1 - \alpha_1 \leq \frac{3}{5}$
			(3, 4, 4, 2)	$\frac{3}{5} \leq 2\beta_1 - \alpha_1 \leq \frac{9}{7}$
			(4, 4, 4, 2)	$2\beta_1 - \alpha_1 \geq \frac{9}{7}$
	g^l	$l_1 = 1$	(4, 8, 8, 4)	$2\beta_1 - \alpha_1 \leq \frac{5}{3}$
			(5, 8, 8, 4)	$\frac{5}{3} \leq 2\beta_1 - \alpha_1 \leq \frac{21}{11}$
			(6, 8, 8, 4)	$2\beta_1 - \alpha_1 \geq \frac{21}{11}$
	g^5	$l_1 = 2$	(8, 12, 12, 8)	
IV. $l_2 > l_3 = l_1 > l_4$	$g^{2\star}$	$l_1 = 1$	(4, 8, 4, 2)	
	$g^{\star+1}$	$l_1 = 2$	(8, 12, 8, 4)	
V. $l_1 = l_2 > l_3 = l_4$	g^1	$l_1 = 1$	(4, 4, 2, 2)	
	g^l	$l_1 = 2$	(8, 8, 4, 4)	
	g^5	$l_1 = 3$	(12, 12, 8, 8)	
VI. $l_3 > l_1 = l_2 = l_4$	$g^{\star}$	$l_1 = 1$	(4, 4, 12, 4)	
VII. $l_1 = l_2 = l_3 > l_4$	$g^{\triangle}$	$l_1 = 2$	(8, 8, 8, 2)	
VIII. $l_1 > l_2 = l_4 > l_3$	$g^{2\star}$	$l_1 = 2$	(8, 4, 2, 4)	
	$g^{\star+1}$	$l_1 = 3$	(12, 8, 4, 8)	
IX. $l_1 > l_2 = l_3 = l_4$	$g^{\star}$	$l_1 = 3$	(12, 4, 4, 4)	

Table 20: Network-contingent equilibrium efforts with one fair player

Network-contingent equilibrium efforts with two fair players

Assume player 1 and player 2 are the fair players. We only consider non-empty networks. Trivially, a player's effort level is fixed to 12 units if he has formed three commitments. It is also easy to show that in any network in which player 1 and player 2 have formed an identical number of commitments and no other player has formed more commitments then both player 1 and player 2 do not move beyond the effort level they committed to. Moreover, no fair player has an incentive to provide an effort beyond the effort level provided by the majority of the other players (2).

CASE I, V, VI, VII, VIII, XI: As player 1 does not move beyond the level he committed to, these cases are identical to the appropriate ones with one fair player (here player 2).

CASE II, X: Players 1 and 2 are the highest committed players and hence stick to the committed level.

CASE III: In any equilibrium in which $e_1 = e_2$ it has to hold that $2\beta_i - \alpha_i \leq \frac{6e_i-9}{2e_i+1}$ and $\beta_i \geq \frac{6e_i-15}{3(2e_i-1)}$, $\forall i \in \{1,2\}$ as otherwise player i has an incentive to increase/decrease his effort. In any equilibrium in which $e_1 > e_2$ it has to hold that $\frac{15-6e_1}{1-2e_1} \leq 2\beta_1 - \alpha_1 \leq \frac{6e_1-9}{2e_1+1}$ and $\frac{5-2e_2}{1-2e_2} \leq \beta_2 \leq \frac{2e_2-3}{2e_2+1}$ as otherwise player 1 (and respectively 2) has an incentive to increase/ decrease his effort.

CASE IV: For $e_1 = e_2$ player 1 and 2 never have an incentive to increase their efforts as $2\alpha_i - \beta_i \geq \frac{9-6e_i}{2e_i+1}$, $\forall i \in \{1,2\}$ is always fulfilled. Hence, in equilibrium it has to hold that $2\beta_i - \alpha_i \geq \frac{15-6e_i}{1-2e_i}$.

CASE IX: In equilibrium it has to hold that $e_1 = e_2$ and that $2\beta_i - \alpha_i \geq \frac{15-6e_i}{1-2e_i}$, $\forall i \in \{1,2\}$ as otherwise player i has an incentive to decrease his effort.

Table 21 and 22 summarize the possible equilibrium effort allocations in all possible non-empty network architectures (accounting for different possible positions of the two fair players). Note again that some of the stated equilibrium efforts are not sustainable with realistic preferences.[41]

Pairwise stable networks with one fair player

We only consider realistic values of (α_i, β_i), $\forall i \in \{1,2,3,4\}$.

- Assume $i \in \{2,3,4\}$ are selfish players. Then if in any network g in which $l_i(g) > 1$, it is easily shown that player i has an incentive to delete a link because the gain from cost reduction overcompensates the loss from the decrease in total effort. Also, it is trivial to show that a selfish player i always prefers one to no link.

- Assume player 1 is the fair player. It can be shown that if $l_1(g) > 1$ and $l_j = 1$, $\forall j \in \{2,3,4\}$ (viz. in $g^{2\star}$ with $l_1 = 2$ and in $g^{\star}$ with $l_1 = 3$), player 1 as an incentive to delete a link. And if g^1 with $l_1 = 0$ player 1 has an incentive to build a link.

Pairwise stable networks with two fair players

We only consider realistic values of (α_i, β_i), $\forall i \in \{1,2,3,4\}$.

- $g^{\emptyset}$ is not pairwise stable as two selfish/two fair players have an incentive to build a link.

- g^1 is not pairwise stable because two selfish/two fair players have an incentive to build a link if they are unconnected.

- $g^{2\star}$ is not pairwise stable because player i with $l_i(g) = 2$ always has an incentive to delete one link irrespective of whether he is selfish or fair.

- g^2 is pairwise stable.

[41] All possible outcomes are generated by relabelling players 1 and 2, and players 3 and 4 respectively. Equilibria are stated as $(e_1^*(g), e_2^*(g), e_3^*(g), e_4^*(g))$. Due to space constraints $l_i(g)$ is abbreviated as l_i.

Case	Network	Position of fair players	Equilibrium	Condition
I. $l_1 > l_3 = l_4 > l_2$	$g^{2\star}$	$l_1 = 2$, $l_2 = 0$	(8, 2, 4, 4)	$\beta_2 \leq \frac{1}{5}$
			(8, 3, 4, 4)	$\frac{1}{5} \leq \beta_2 \leq \frac{3}{7}$
			(8, 4, 4, 4)	$\beta_2 \geq \frac{3}{7}$
	$g^{\star+1}$	$l_1 = 3$, $l_2 = 1$	(12, 4, 8, 8)	$\beta_2 \leq \frac{5}{9}$
			(12, 5, 8, 8)	$\frac{5}{9} \leq \beta_2 \leq \frac{7}{11}$
			(12, 6, 8, 8)	$\frac{7}{11} \leq \beta_2 \leq \frac{9}{13}$
			(12, 7, 8, 8)	$\frac{9}{13} \leq \beta_2 \leq \frac{11}{15}$
			(12, 8, 8, 8)	$\beta_2 \geq \frac{11}{15}$
II. $l_1 = l_2 > l_3 = l_4$	g^1	$l_1 = l_2 = 1$	(4, 4, 2, 2)	
	g^l	$l_1 = l_2 = 2$	(8, 8, 4, 4)	
	g^5	$l_1 = l_2 = 3$	(12, 12, 8, 8)	
III. $l_3 = l_4 > l_1 = l_2$	g^1	$l_1 = l_2 = 0$	(2, 2, 4, 4)	$2\beta_i - \alpha_i \leq \frac{3}{5}$, $\forall i \in \{1, 2\}$
			(3, 3, 4, 4)	$2\beta_i - \alpha_i \leq \frac{9}{7}$ and $\beta_i \geq \frac{1}{5}$, $\forall i \in \{1, 2\}$
			(4, 4, 4, 4)	$2\beta_i - \alpha_i \geq \frac{9}{7}$ and $\beta_i \geq \frac{3}{7}$, $\forall i \in \{1, 2\}$
			(3, 2, 4, 4)	$\frac{3}{5} \leq 2\beta_1 - \alpha_1 \leq \frac{9}{7}$ and $\beta_2 \leq \frac{1}{5}$
			(4, 2, 4, 4)	$\frac{9}{7} \leq 2\beta_1 - \alpha_1 \leq \frac{5}{3}$ and $\beta_2 \leq \frac{1}{5}$
			(4, 3, 4, 4)	$\frac{9}{7} \leq 2\beta_1 - \alpha_1 \leq \frac{5}{3}$ and $\frac{1}{5} \leq \beta_2 \leq \frac{3}{7}$
	g^l	$l_1 = l_2 = 1$	(4, 4, 8, 8)	$2\beta_i - \alpha_i \leq \frac{5}{3}$, $\forall i \in \{1, 2\}$
			(5, 5, 8, 8)	$2\beta_i - \alpha_i \leq \frac{21}{11}$ and $\beta_i \geq \frac{5}{9}$, $\forall i \in \{1, 2\}$
			(6, 6, 8, 8)	$\beta_i \geq \frac{7}{11}$, $\forall i \in \{1, 2\}$
			(7, 7, 8, 8)	$\beta_i \geq \frac{9}{13}$, $\forall i \in \{1, 2\}$
			(8, 8, 8, 8)	$\beta_i \geq \frac{11}{15}$, $\forall i \in \{1, 2\}$
			(5, 4, 8, 8)	$\frac{5}{3} \leq 2\beta_1 - \alpha_1 \leq \frac{21}{11}$ and $\beta_2 \leq \frac{5}{9}$
			(6, 4, 8, 8)	$\frac{21}{11} \leq 2\beta_1 - \alpha_1 \leq \frac{27}{13}$ and $\beta_2 \leq \frac{5}{9}$
			(6, 5, 8, 8)	$\frac{21}{11} \leq 2\beta_1 - \alpha_1 \leq \frac{27}{13}$ and $\frac{5}{9} \leq \beta_2 \leq \frac{7}{11}$
	g^5	$l_1 = l_2 = 2$	(8, 8, 12, 12)	
IV. $l_3 > l_1 = l_2 > l_4$	$g^{2\star}$	$l_1 = l_2 = 1$	(4, 4, 8, 2)	
			(5, 5, 8, 2)	$2\beta_i - \alpha_i \geq \frac{5}{3}$, $\forall i \in \{1, 2\}$
			(6, 6, 8, 2)	$2\beta_i - \alpha_i \geq \frac{21}{11}$, $\forall i \in \{1, 2\}$
	$g^{\star+1}$	$l_1 = l_2 = 2$	(8, 8, 12, 4)	

Table 21: Network-contingent equilibrium efforts with two fair players, Part I

- $g^\star$ is not pairwise stable as player i with $l_i = 3$ always has an incentive to delete one link irrespective of whether he is selfish or fair.

- g^l is not pairwise stable because any player i with $l_i(g) = 2$ has an incentive to delete one link irrespective of whether he is selfish or fair.

- $g^\triangle$, g°, $g^{\star+1}$, g^5, g^N are not pairwise stable because there is always at least on selfish player i with $l_i(g) > 1$ who has an incentive to delete one link.

Case	Network	Position of fair players	Equilibrium	Condition
V. $l_1 = l_3 > l_2 = l_4$	g^1	$l_1 = 1$, $l_2 = 0$	(4, 2, 4, 2)	$2\beta_2 - \alpha_2 \leq \frac{3}{5}$
			(4, 3, 4, 2)	$\frac{3}{5} \leq 2\beta_2 - \alpha_2 \leq \frac{9}{7}$
			(4, 4, 4, 2)	$2\beta_2 - \alpha_2 \geq \frac{9}{7}$
	g^l	$l_1 = 2$, $l_2 = 1$	(8, 4, 8, 4)	$2\beta_2 - \alpha_2 \leq \frac{5}{3}$
			(8, 5, 8, 4)	$\frac{5}{3} \leq 2\beta_2 - \alpha_2 \leq \frac{21}{11}$
			(8, 6, 8, 4)	$2\beta_2 - \alpha_2 \geq \frac{21}{11}$
	g^5	$l_1 = 3$, $l_2 = 2$	(12, 8, 12, 8)	
VI. $l_1 > l_2 = l_4 > l_3$	$g^{2\star}$	$l_1 = 2$, $l_2 = 1$	(8, 4, 2, 4)	
	$g^{\star+1}$	$l_1 = 3$, $l_2 = 2$	(12, 8, 4, 8)	
VII. $l_4 > l_3 = l_1 > l_2$	$g^{2\star}$	$l_1 = 1$, $l_2 = 0$	(4, 2, 4, 8)	$\beta_2 \leq \frac{1}{5}$
			(4, 3, 4, 8)	$\frac{1}{5} \leq \beta_2 \leq \frac{3}{7}$
			(4, 4, 4, 8)	$\beta_2 \geq \frac{3}{7}$
	$g^{\star+1}$	$l_1 = 2$, $l_2 = 1$	(8, 4, 8, 12)	$\beta_2 \leq \frac{5}{9}$
			(8, 5, 8, 12)	$\frac{5}{9} \leq \beta_2 \leq \frac{7}{11}$
			(8, 6, 8, 12)	$\frac{7}{11} \leq \beta_2 \leq \frac{9}{13}$
			(8, 7, 8, 12)	$\frac{9}{13} \leq \beta_2 \leq \frac{11}{15}$
			(8, 8, 8, 12)	$\beta_2 \geq \frac{11}{15}$
VIII. $l_1 > l_2 = l_3 = l_4$	$g^\star$	$l_1 = 3$, $l_2 = 1$	(12, 4, 4, 4)	
IX. $l_4 > l_1 = l_2 = l_3$	$g^\star$	$l_1 = l_2 = 1$	(4, 4, 4, 12)	
			(5, 5, 4, 12)	$2\beta_i - \alpha_i \geq \frac{5}{3}$, $\forall i \in \{1, 2\}$
X. $l_1 = l_2 = l_4 > l_3$	$g^\triangle$	$l_1 = l_2 = 2$	(8, 8, 2, 8)	
XI. $l_1 = l_3 = l_4 > l_2$	$g^\triangle$	$l_1 = 2$, $l_2 = 0$	(8, 2, 8, 8)	$\beta_2 \leq \frac{1}{5}$
			(8, 3, 8, 8)	$\frac{1}{5} \leq \beta_2 \leq \frac{3}{7}$
			(8, 4, 8, 8)	$\frac{3}{7} \leq \beta_2 \leq \frac{5}{9}$
			(8, 5, 8, 8)	$\frac{5}{9} \leq \beta_2 \leq \frac{7}{11}$
			(8, 6, 8, 8)	$\frac{7}{11} \leq \beta_2 \leq \frac{9}{13}$
			(8, 7, 8, 8)	$\frac{9}{13} \leq \beta_2 \leq \frac{11}{15}$
			(8, 8, 8, 8)	$\beta_2 \geq \frac{11}{15}$
XII. $l_1 = l_2 = l_3 = l_4$	g^2	$l_1 = l_2 = 1$	(4, 4, 4, 4)	
	g°	$l_1 = l_2 = 2$	(8, 8, 8, 8)	
	g^N	$l_1 = l_2 = 3$	(12, 12, 12, 12)	

Table 22: Network-contingent equilibrium efforts with two fair players, Part II

Pairwise stable networks with three fair player

We only consider networks with $l(g) \geq 3$ and realistic values of (α_i, β_i), $\forall i \in \{1, 2, 3, 4\}$.

- $g^\triangle$, g^l are not pairwise stable as any player i with $l_i = 2$ has an incentive to delete one link irrespective of whether he is selfish or fair.

- g°, g^5, g^N are not pairwise stable as there is always at least one selfish player i with $l_i > 1$ who has an incentive to delete one link.

- $g^{\star+1}$, $g^\star$ are not pairwise stable as player i with $l_i = 3$ has an incentive to delete one link irrespective of whether he is selfish or fair.

Pairwise stable networks with four fair players

We only consider networks with $l(g) \geq 3$ and realistic values of (α_i, β_i), $\forall i \in \{1, 2, 3, 4\}$.

- $g^\triangle$, g^l are not pairwise stable as any player i with $l_i = 2$ has an incentive to delete one link.
- $g^\star$, $g^{\star+1}$, g^5 are not pairwise stable as player i with $l_i = 3$ has an incentive to delete one link.
- g^N is only pairwise stable for $\beta_i \geq 0.9$ which is not a realistic value.
- g° can be pairwise stable for the denoted range in which no player i has an incentive to delete a link.

This completes the proof. □

References

Andreoni, J., and H. Varian (1999): "Preplay contracting in the Prisoner's Dilemma," *Proceedings of the National Academy of Sciences*, 96, 10933–10938.

Aumann, R. J., and R. B. Myerson (1988): "Endogenous Formation of Links Between Players and Coalitions: An Application of the Shapley Value," in *The Shapley Value*, ed. by A. Roth, pp. 175–191. Cambridge University Press.

Bochet, O., T. Page, and L. Putterman (forthcoming): "Communication and Punishment in Voluntary Contribution Experiments," *Journal of Economic Behavior and Organization.*

Cason, T. N., and F. U. Khan (1999): "A laboratory study of voluntary public goods provision with imperfect monitoring and communication," *Journal of Development Economics*, 58, 533–552.

Charness, G., G. Fréchette, and C.-Z. Qin (2005): "Endogenous Transfers in the Prisoner's Dilemma Game: An Experimental Test of Cooperation and Coordination," mimeo.

Charness, G., and M. Rabin (2002): "Understanding Social Preferences with Simple Tests," *Quarterly Journal of Economics*, 117, 817–869.

Coase, R. H. (1960): "The Problem of Social Cost," *The Journal of Law and Economics*, 3, 1–44.

Croson, R. (1998): "Feedback in voluntary contribution mechanisms: an experiment in team production," in *Research in Experimental Economics*, ed. by R. M. Isaac. Greenwich: JAI Press.

Dixit, A., and M. Olson (2000): "Does voluntary participation undermine the Coase Theorem?," *Journal of Public Economics*, 76, 309–335.

DUFWENBERG, M., AND G. KIRCHSTEIGER (2004): “A Theory of Sequential Reciprocity,” *Games and Economic Behavior*, 47, 268–298.

FAFCHAMPS, M. (1992): “Solidarity Networks in Preindustrial Societies: Rational Peasants with a Moral Economy,” *Economic Development and Cultural Change*, 41, 147–174.

FAFCHAMPS, M., AND S. LUND (2003): “Risk Sharing Networks in Rural Philippines,” *Journal of Development Economics*, 71, 261–287.

FALK, A., AND U. FISCHBACHER (forthcoming): “A theory of reciprocity,” *Games and Economic Behavior*.

FALKINGER, J., E. FEHR, S. GÄCHTER, AND R. WINTER-EBMER (2000): “A Simple Mechanism for the Efficient Provision of Public Goods: Experimental Evidence,” *American Economic Review*, 90, 247–264.

FEHR, E., AND S. GÄCHTER (2000): “Cooperation and Punishment,” *American Economic Review*, 90, 980–994.

——— (2002): “Altruistic Punishment in Humans,” *Nature*, 415, 137–140.

FEHR, E., AND K. M. SCHMIDT (1999): “A Theory of Fairness, Competition, and Cooperation,” *Quarterly Journal of Economics*, 114, 817–868.

——— (2003): “Theories of Fairness and Reciprocity: Evidence and Economic Applications,” in *Advances in Economic Theory, Eigth World Congress of the Econometric Society, Vol. I*, ed. by S. T. M. Dewatripont, and L. P. Hansen, pp. 208–257. Cambridge: Cambridge University Press.

FISCHBACHER, U., S. GÄCHTER, AND E. FEHR (2001): “Are people conditionally cooperative? Evidence from a public goods experiment,” *Economics Letters*, 71, 397–404.

GONZÁLEZ, L. G., G. GONZÁLEZ-FARÌAS, AND M. V. LEVATI (2005): “Logit estimation of conditional cooperation in a repeated public goods experiment,” Discussion Papers on Strategic Interaction 5-2005, Max Planck Institute for Research into Economic Systems.

GOREN, H., R. KURZBAN, AND A. RAPOPORT (2003): “Social loafing vs. social enhancement: Public goods provisioning in real-time with irrevocable commitments,” *Organizational Behavior and Human Decision Processes*, 90, 277–290.

GREINER, B. (2003): “An Online Recruitment System for Economic Experiments,” in *Forschung und wissenschaftliches Rechnen. GWDG Bericht 63*, ed. by K. Kremer, pp. 79–93. Göttingen: Ges. für Wiss. Datenverarbeitung.

HUME, D. (1928): “Book III of Morals,” in *A Treatise of Human Nature*, ed. by L. A. Selby-Bigge. Oxford: Clarendon Press.

ISAAC, R. M., J. M. WALKER, AND S. H. THOMAS (1984): “Divergent evidence on free riding: An experimental examination of possible explanations,” *Public Choice*, 43, 113–149.

Jackson, M. O., and A. Wolinsky (1996): "A Strategic Model of Social and Economic Networks," *Journal of Economic Theory*, 71, 44–74.

Kurzban, R., K. McCabe, V. L. Smith, and B. J. Wilson (2001): "Incremental commitment and reciprocity in a real time public goods game," *Personality and Social Psychology Bulletin*, 27, 1662–1673.

Montgomery, M. R., and R. Bean (1999): "Market Failure, Government Failure, and the Private Supply of Public Goods: The Case of Climate-Controlled Walkway Networks," *Public Choice*, 99, 403–437.

Murgai, R., P. Winters, E. Sadoulet, and A. de Janvry (2002): "Localized and Incomplete Mutual Insurance," *Journal of Development Economics*, 67, 245–274.

Normann, H.-T., and B. Wallace (2005): "The Impact of the Termination Rule on Cooperation in a Prisoner's Dilemma Experiment," mimeo.

Ostrom, E. (1990): *Governing the Commons: The Evolution of Institutions for Collective Action.* Cambridge University Press, Cambridge.

——— (2000): "Collective Action and the Evolution of Social Norms," *Journal of Economic Perspectives*, 14, 137–158.

Rabin, M. (1993): "Incorporating Fairness into Game Theory and Economics," *American Economic Review*, 83, 1281–1302.

Ray, D., and R. Vohra (2001): "Coalitional Power and Public Goods," *Journal of Political Economy*, 109, 1355–1384.

Sell, J., and R. K. Wilson (1991): "Levels of Information and Contributions to Public Goods," *Social Forces*, 70, 107–124.

Sobel, J. (2004): "Interdependent Preferences and Reciprocity," mimeo.

Tang, S. Y. (1992): *Institutions and Collective Action: Self-Governance in Irrigation.* Institute of Contemporary Studies Press, San Francisco.

Varian, H. R. (1994): "A Solution to the Problem of Externalities When Agents Are Well-Informed," *American Economic Review*, 84, 1278–1293.

Wade, R. (1988): *Village Republics: Economic conditions for collective action in South India.* Cambridge University Press, Cambridge.

Watts, A. (2002): "Non-myopic formation of circle networks," *Economics Letters*, 74, 277–282.

Weimann, J. (1994): "Individual behavior in a free-riding experiment," *Journal of Public Economics*, 54, 185–200.

Wilson, R. K., and J. Sell (1997): ""Liar, Liar ...": Cheap Talk and Reputation in Repeated Public Goods Settings," *Journal of Conflict Resolution*, 41, 695–717.