Engelmann, Dirk; Grimm, Veronika

Working Paper

Mechanisms for efficient voting with private information about preferences

IWQW discussion paper series, No. 03/2008

Provided in cooperation with:
Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU)

IWQW

Institut für Wirtschaftspolitik und Quantitative Wirtschaftsforschung

Diskussionspapier
Discussion Papers

No. 03/2008

Mechanisms for Efficient Voting with Private Information about Preferences

Dirk Engelmann
Royal Holloway, University of London

Veronika Grimm
University of Erlangen-Nuremberg

ISSN 1867-6707

Mechanisms for Efficient Voting with Private Information about Preferences.*

Dirk Engelmann†
Royal Holloway, University of London

Veronika Grimm‡
University of Erlangen–Nuremberg

August 11, 2008

Abstract

We experimentally study behavior in a simple voting game where players have private information about their preferences. With random matching, subjects overwhelmingly follow the dominant strategy to exaggerate their preferences, which leads to inefficiency. We analyze an exogenous linking mechanism suggested by Jackson and Sonnenschein (2007) as well as repeated interaction in different settings, which could allow endogenous linking mechanisms to evolve. We find that applying the exogenous mechanism captures nearly all achievable efficiency gains, whereas repeated interaction leads to significant gains in truthful representation and efficiency only in a setting where players can choose their partners.

Keywords: Experimental Economics, Mechanism Design, Implementation, Linking, Bayesian Equilibrium, Efficiency.
JEL classification: A13, C72, C91, C92, D64, D72, D80.

*We thank Todd Kaplan, Wieland Müller, Hans Normann and Aljaž Ule as well as seminar participants at the University of Alicante, Royal Holloway, Tinbergen Institute, the University of Hannover, the University of Magdeburg, the University of Karlsruhe, Dundee University, the University of St. Andrews, the University of Copenhagen, the University of Exeter, as well as participants at the ESA meetings in Montreal and Alessandria, the EEA congress in Vienna and the Verein für Socialpolitik meeting in Bayreuth for helpful comments and suggestions.

†Department of Economics, Royal Holloway, University of London, Egham, Surrey, TW20 0EX, United Kingdom, dirk.engelmann@rhul.ac.uk

‡University of Erlangen–Nuremberg, Department of Economics, Lange Gasse 20, D–90403 Nürnberg, Germany, veronika.grimm@wiso.uni-erlangen.de

1 Introduction

When agents have private information about their preferences, there is generally no incentive compatible mechanism that implements the ex-ante socially efficient solution.[1] In a recent paper Jackson and Sonnenschein (2007) readdress this issue. They demonstrate how the limitations that incentive constraints impose on the attainment of socially desirable outcomes can be overcome when independent social decision problems are linked. They propose to "ration" or "budget" the agents' representations in accordance with their empirical distribution. Imposing those budgets increases efficiency of the outcomes as compared to deciding on each problem separately since it allows to ask the players the question: "which decision do you care more about?" Let us illustrate this point with an example.

Consider coalition talks between two political parties. In the course of the negotiations, agreements on a variety of topics have to be reached. From a social point of view (assuming that the parties reflect the preferences of their voters), for each single decision it would be desirable if the party succeeded that cares more about the issue. Assume that it is publicly known that each party is equally likely to care a lot, or a little about each single issue on the table.[2] If, however, only the parties themselves know their preferences exactly, they have an incentive to pretend to assign high importance to every single issue in order to affect the result of the negotiations in their favor. A budget in the sense of Jackson and Sonnenschein would restrict both sides to state a high importance for only half of the issues that are negotiated. The parties then have an incentive to utilize their budget first on issues that are indeed important to them, until the budget is exhausted.

The example also demonstrates an obvious difficulty that arises upon implementation. Institutions would be needed in order to implement and enforce a budget. While two parties involved in coalition talks might ex–ante agree on procedures that effectively result in budgeting, in many situations, such institutions do not exist and are difficult to establish. Moreover, players do not only have an incentive to lie concerning their types, but concerning their distribution of types: By claiming that they care a lot about almost everything that is

[1]See the well known contribution by Myerson and Satterthwaite (1983).

[2]This is, of course, oversimplified but it facilitates to make our point.

being decided they could achieve a more favorable budget, which complicates the ex–ante agreement on institutions that enable budgeting.

However, in practice economic and political players are usually aware that they can only exploit benefits (i.e. efficiency gains) in a stable relationship if they do not upset their counterpart by overweighing their own interest. Thus, social interaction could lead to endogenous budgets that are enforced by the threat of retaliation. One of the problems that come with the induced endogenous budget is that perception of compliance to such a virtual budget need not be the same among the agents involved in the interaction, as we will discuss below.

Following our above arguments that (i) a social planner who could enforce budgets is not always available, but that (ii) budgets could also arise endogenously, in this paper we compare the effectiveness of exogenous budgets as proposed by Jackson and Sonnenschein with various forms of social interaction that have the potential to imply an endogenous budget. In particular, we study *stable partnerships*, *reputation building* and *competition for partners*. We do this in the context of an experiment where two players engage in a fictitious joint project. They always disagree on the version of the project to be implemented, but the intensity of their preferences is private information. The agents decide on the version to be implemented in a simple voting game where they have to indicate the intensity of their preferences. The version which is preferred strongest is chosen (in case of a tie, a coin flip decides). We vary the matching protocol and whether subjects face an exogenous budget on preference representation or not. We find that exogenous budgets help players to reap almost all achievable efficiency gains. Among the social interaction treatments, only competition for partners leads to a significant increase in truthful representation of preferences and efficiency. Two control treatments serve to assess possible explanations for the relatively low effectiveness of social interaction. The ambiguity of signals does not appear to be crucial. In contrast, the coordination problem in the sense that all involved players should understand how to reap the efficiency gains and need to implicitly agree on a budget, seems to be of major importance. Competition for partners enables players to reduce this problem.

Our paper is related to three areas in recent literature. The paper by Jackson and Sonnenschein was inspired and generalizes the storable votes idea of Casella (2005). Casella, Gelman and Palfrey (2006) study the storable votes mechanism experimentally and find that players make effective use of the opportunity to store votes. Even though equilibrium strategies are difficult to compute, realized efficiency levels are very close to the theoretical prediction. Hortala–Vallve (2004) generalizes the storable votes mechanism to "qualitative voting", which allows players to freely allocate votes across decisions. He also assumes that they are informed about the intensity of their preferences concerning all decisions from the start. Hortala–Vallve and Llorente–Saguer (2006) present experimental support. Subjects generally vote in accordance with the equilibrium predictions if they vote over two issues. If players vote over more than two issues, they deviate more frequently from the qualitative voting equilibrium but still reach efficiency close to the equilibrium level. These results are well in line with our result on the effectiveness of the Jackson–Sonnenschein mechanism. Kaplan and Ruffle (2005) study a market–entry game with private information. In contrast to our results, they find that players coordinate well on efficient cut-off strategies. The effects of competition for partners in trust games is studied by Huck, Lünser, and Tyran (2006). In line with our results, they find that competition increases trustworthiness beyond the level achieved through reputation building alone.[3] A theoretical analysis of partner choice in dilemma games is provided by Ule (2006).

The paper is organized as follows: In Section 2 we introduce the game we study, discuss our main research questions in more detail, develop the experimental design, and introduce some concepts we use to evaluate our data. Section 3 presents a theoretical analysis of the different settings and states our main hypotheses. In Section 4 we report the results. Section 5 is devoted to the question which factors drive successful cooperation in the absence of exogenous budgeting. Section 6 concludes.

2 An Experiment on Linking Decisions

In Section 2.1 we present a slightly modified version of one of Jackson and Sonnenschein's (2007) examples (in order to illustrate their point), and introduce two experimental treat-

[3]Coricelli, Fehr, and Fellner (2003) study partner selection in public good experiments. Interestingly, the contribution levels are highest for unidirectional partner selection. Hauk and Nagel (2001) find similar results in a prisoner's dilemma experiment.

ments that shall evaluate the empirical relevance of the incentive problem and the effectiveness of the Jackson–Sonnenschein mechanism. Then, in Section 2.2, we argue that also social interaction might solve the problem by implementing an endogenous budget. We illustrate that players have an incentive to cooperate in order to realize (and share) efficiency gains and propose three treatments that imply different incentives to do so.

2.1 The Idea of Exogenous Budgets

Suppose that two players, a and b, are engaged in a joint project. It is common knowledge that the players always disagree on the version of the project to be chosen. Let us call the version preferred by player a version a, and the version preferred by player b version b. Each player receives a positive payoff only if his preferred version of the project is chosen. A player's *intensity* of preference for his preferred version, however, is private information. The intensity can either be strong (s) or weak (w), where $s > w$. Both cases are equally likely.[4]

Now suppose that a social planner wants to choose the version of the project that maximizes the sum of the utilities. If the intensity of preferences is the same for both players, the social planner is indifferent which version to choose and can flip a coin. Otherwise, he wants to choose the version preferred by the player with the stronger intensity of preference. Table 1 illustrates this social choice function.

		player a's preference	
		w	s
player b's preference	w	coin	a
	s	b	coin

Table 1: Efficient social choice function

The problem with this social choice function is, however, that it is not incentive compatible. That is, if it is applied to the players' *stated* preferences, it is each player's dominant

[4]The intensity of preferences corresponds to the payoff received if the agent's preferred version is chosen. In the experiment we consider the case that $s = 2w$.

strategy to always state a strong preference, whatever preference he observed. The closest social choice function that can be implemented through an incentive compatible mechanism is illustrated in Table 2.

player a states

		w	s
player b states	w	coin	coin
	s	coin	coin

Table 2: Incentive Compatible Mechanism

Note that under this social choice function the version is always determined by a flip of a coin, independently of the preferences stated by the two players.[5] This social choice function is ex–post Pareto efficient, but not ex–ante. The reason is that once players have observed their stated preferences it is not possible to write a contract that improves the situation of both players. Ex–ante, however, this is possible. The linking mechanism exploits this fact.

Linking two independent decisions Now consider the case that the two players have to decide on two independent problems simultaneously. First, note that if players separately vote over the two problems, what was ex–ante Pareto inefficiency in the single decision problem becomes ex–post inefficiency in the situation where players decide simultaneously on two problems. To see this, consider the case that each player has one strong and one weak preference and that player a's preference is strong for the first project whereas player b's preference is strong for the second one. Now, if for the first project player b's version is chosen and for the second project player a's, then, even ex–post, players would benefit from turning around the decision.

Jackson and Sonnenschein propose the following mechanism that links the two problems. When stating preference intensities for the two (independent) projects, each player is

[5] In case both players might prefer the same version with some positive probability, the mechanism would be extended to simply voting on the version to be executed, without taking into account the intensities of preferences and flipping a coin whenever the agents disagree. In the experiment we only consider the case where the players disagree.

allowed to state a strong preference only once. The ex–ante efficient social choice function is then applied to the constrained announcements. Jackson and Sonnenschein show that there is a Bayesian equilibrium of their mechanism with the following features:

- If an agent's intensity of preference differs across the two problems then he or she announces truthfully.
- If an agent has two preference intensities of the same magnitude, then the agent randomly chooses which problem to announce the strong preference for.

Although the equilibria of the linked mechanism are not Pareto efficient (neither ex–ante nor ex–post), the equilibrium outcomes still Pareto dominate from any perspective (ex–ante, interim, or ex–post) voting on the problems separately. The reason is that linking two problems allows to ask the players the question "Which decision do you care more about?" Jackson and Sonnenschein (2007) show that linking more decisions helps further, and in the limit their mechanism leads to full Pareto efficiency.

As a first step of our experimental study, we establish that the problem analyzed by Jackson and Sonnenschein is indeed empirically relevant, i.e. experimental subjects follow the incentives to overstate their preferences, and investigate whether it can be satisfactorily solved by the mechanism they propose. For this purpose we run two treatments:

Treatment I: Random Matching (RAN) In this treatment, we test whether in an environment without any incentive to be honest, honesty indeed breaks down. Subjects are randomly rematched in pairs within a relatively large group (8 subjects) without any opportunity for identification and hence no opportunity to build a reputation for honesty that could either be reciprocated or attract new partners. In each period, a pair of subjects is engaged in one fictitious project as described above. Intensities of preferences are randomly drawn, independently among subjects and periods. After subjects are informed about the intensity of their preference, they state a preference and then the version supported by the stronger stated preference is implemented, i.e. this subject receives a payoff while the other receives nothing. The version is chosen randomly if the stated intensities are equal. At the end of the period, each player is informed about both stated preferences, the version that is implemented and his own resulting payoffs.

Treatment II: Exogenous Budgets (EXO) In this treatment subjects were also rematched randomly, as in the previous treatment. However, each subject faced a budget corresponding to the expected distribution of preferences, that is, he could state at most 20 strong preferences over the 40 periods that the experiment lasted.[6]

2.2 Can Budgets Arise Endogenously?

Different forms of social interaction that allow for the formation of long–term partnerships or reputation building may — to a different extent — be capable of promoting cooperation among players by endogenously creating a need to budget stated preferences. The second aim of our study is to address this issue. We run several additional treatments to investigate whether subjects manage to realize potential efficiency gains without exogenous budgeting if (a) they interact repeatedly, (b) the environment offers the chance to build a reputation, or (c) they have to compete for partners.

Note that under incomplete information about preferences budgets are needed to establish cooperation, even if they are not enforced exogenously. Since a player cannot assess the other's honesty directly, any conditionally cooperative strategy can only be based on the distribution of the other's stated preferences. Hence, if one player follows a conditionally cooperative strategy, this implies that the other player needs to budget his stated preferences, independent of his real preferences.[7] Obviously, it is required that decisions are sequential for a player to be able to reciprocate violations of the endogenous budget,[8] whereas an exogenous budget can be applied to several decisions that are made simultaneously as well

[6]We did not limit the number of weak preferences that could be stated, as a rational player should always exploit his budget for strong preferences completely. Indeed, almost all subjects did.

[7]A possible conditionally cooperative strategy that players might follow would be "stochastic tit–for–tat", i.e. switching to stating always strong preferences in the next m periods if the other player has stated more than $n < m$ strong preferences in the last m periods. Another possible, highly sophisticated strategy would be "binomial trigger", i.e. switching to always stating a strong preference once a binomial test applied to the other player's sequence of stated preferences allows one to reject the hypothesis that this sequence is random at some pre–determined level.

[8]To be more precise, several decisions can be made simultaneously as long as these are repeated.

as to those made sequentially. Another problem that arises with endogenous budgets is that a player needs to know the other player's strategy in order to know his own budget.[9]

We consider the following treatments.

Stable Partnerships (FIX) In this treatment, each pair stays together for the whole course of the experiment. By comparing this to the random matching treatment (RAN), we can assess to what degree subjects are able to realize the mutual gains from honesty in a long–term relationship. Reciprocating honesty with honesty increases the expected payoffs for both subjects (where honesty can only stochastically be detected via an endogenous budget). In order to facilitate keeping track of past decisions, after each period players observe a summary of the history of the past periods played with their partner. In particular, they observe announced preferences and the decisions that were taken within their pair in all preceding periods.

Random Link Formation in Stable Groups: The Scope for Reputation Building (RLK) In this treatment, subjects interact in fixed groups of four, while partnerships are still only formed by pairs. In each period, from each subject one link is established to another, randomly selected subject in the group of four. Each of these links corresponds to one project. Hence each subject can in any particular period be involved in one to four projects which are independent in terms of valuations and implementation. That is, for each of the projects in which a subject is involved, his or her valuation is independently drawn. If there is a link from subject 1 to subject 2 and a link from subject 2 to subject 1, these are two independent projects. After being informed about all the projects to be executed in their group and about their respective valuations for each of the projects they are involved in, all subjects simultaneously state their preferences for all projects they are involved in.

[9] On the other hand, recall that exogenous budgeting forces the agents to lie if their true distribution of preference intensities does not perfectly match the underlying distribution. Thus, since endogenous budgets can have more flexibility, it is in principle even possible that they reach higher efficiency than exogenous ones.

At the end of each period all subjects are informed about all stated preferences and implemented project versions in their group. Then they are shown a screen with the history of all stated preferences of all players in their group. This treatment allows for reputation building in a more complicated setting than a simple fixed pairing. In particular, in addition to direct reciprocation as in the stable partnerships treatment, this treatment allows also for indirect reciprocity and strategic reputation building.[10] The specific design of this treatment is necessary to serve as a benchmark for our next treatment.

Voluntary Link Formation in Stable Groups: Competition for Partnerships (CMP) This treatment differs from the random link formation treatment only in that the link originating from each subject is not randomly chosen, but is chosen by the subject. That is, at the first stage of every period, each subject chooses one of the other three subjects as a partner for one project. Then, as in the random link treatment, all subjects are informed about all links and about their (independently and randomly chosen) preferences concerning the projects they are involved in. They then choose simultaneously their stated preferences, and implementation and feed-back is as above in RLK.

By comparing the behavior in this treatment with that in the random link treatment, we can assess the impact that the competition for partners has on top of the incentives for reputation building. Being involved in more projects is beneficial because the expected payoff from each single project is nonnegative. This incentive to increase the number of partnerships could actually result in subjects being even nicer than truthful and trying to build a reputation of almost always giving in, which in turn could lead to inefficiencies.

2.3 Further Details of the Experimental Implementation

In all our treatments, pairs of players had to decide on a joint project as it has been described in Section 2.1. In the experiment, a weak (strong) preference corresponded to a payoff of 30 (60) Pence if the desired version of the project was chosen. In each period,

[10] See Engelmann and Fischbacher (2003) for experimental evidence that many experimental subjects are indirectly reciprocal and that they also recognize the incentives for strategic reputation building in an environment where indirect reciprocity is possible.

for each of a subject's decisions (remember that one subject might have been involved in more than one decision per period), the intensity of preference was drawn randomly and independently across decisions, periods, and subjects, where each possible intensity (30 or 60) was equally likely. In each treatment, 40 periods were played. The payoff was counted directly in UK Pence. At the end of the experiment, the earnings were paid in cash in Pound Sterling.

All experimental sessions were computerized using z-Tree (Fischbacher, 2007) and were conducted in the experimental laboratory at Royal Holloway. In each experimental session, 8 to 16 subjects participated. The total number of subjects was 148 (including two control treatments discussed below). We conducted one session for the treatment with fixed pairs and two for each of the other treatments. See Table 3 for details.

treatment	# subjects in sessions	# independent obs. (# sub/obs))
Random Matching RAN	16, 8	3(8)
Exogenous Budgets EXO	16, 8	3(8)
Stable Partnerships FIX	16	8(2)
Random Links RLK	16, 12	7(4)
Competition CMP	12, 16	7(4)

Table 3: Number of subjects, independent observations and subjects per independent observation for the different experimental treatments

Written instructions were distributed at the beginning of the experiment and subjects could go through them at their own pace. After subjects had answered a set of control questions, the key features of the experiment were orally summarized by one of the experimenters (the same in all sessions). The experiments took between 45 and 120 minutes (including reading the instructions, answering a post-experimental questionnaire and receiving payments).[11] Average earnings ranged from 9.01 (treatment FXI explained below) to 18.64 (competition treatment) with an overall average of 13.13.[12]

[11]There are remarkable differences in the amount of time the experiment itself took. It ranged from 15-20 minutes in the random matching treatment to 60-80 minutes in the competition treatment.

[12]Note that the large differences in average payoffs are not due to much more successful cooperation, but

2.4 Measures of Behavior

We now discuss several measures of behavior that help us to assess whether endogenous budgets arise and that allow us to analyze the extent to which subjects manage to overcome incentive constraints. Explicit hypotheses that we state in the next section will refer to these measures. There are two important aspects of behavior that we can compare between treatments. First, how do the treatment variations affect how honestly subjects represent their preferences? Second, does this translate into differences in efficiency?

Honesty Rates. We measure how truthfully subjects state their preferences separately for the case that their true preference is strong or weak. The measures are

$$H_s = \frac{\#\text{truthfully stated strong preferences}}{\#\text{true strong preferences}} \tag{1}$$

$$H_w = \frac{\#\text{truthfully stated weak preferences}}{\#\text{true weak preferences}} \tag{2}$$

Efficiency. The efficiency is measured in expected terms with respect to the random draws in case of equal stated preferences.

Denote by i and j the two players involved in a project p. Denote the maximum achievable surplus in project p by $S_p^{\max} = \max\{\text{RealPref}_p^i, \text{RealPref}_p^j\}$ and the minimum achievable surplus by $S_p^{\min} = \min\{\text{RealPref}_p^i, \text{RealPref}_p^j\}$, where RealPref_p^k is the payoff of player k, $k = i, j$, if his preferred version of the project is chosen, i.e. the true intensity of his preference. Furthermore, $S_p^{\text{real}} = \text{RealPref}_p^i * \text{Win}_p^i + \text{RealPref}_p^j * (1 - \text{Win}_p^i)$ denotes the in project p actually realized surplus, where Win_p^i is a dummy that is 1 if the preferred version of player i has been chosen. If both players state different preferences, S_p^{real} is just equal to the preference of the player who stated the stronger preference.

In case of equal stated preferences, whose preferred version of the project will be chosen is determined by a random draw. Since we do not want our measure of efficiency to be influenced by the outcome of this random draw, we consider the expected achieved surplus

primarily occur because there are twice as many projects per subject in RLK and CMP as in the other treatments.

(given the preferences drawn and the behavior in the experiment but taking expectations with respect to the allocation),

$$E[S_p^{\text{real}}] = \text{Equal}_p * \left(\frac{1}{2}\text{RealPref}_p^i + \frac{1}{2}\text{RealPref}_p^j\right) + (1 - \text{Equal}_p) * S_p^{\text{real}}, \tag{3}$$

where Equal_p is a dummy that is 1 if both players state equal preferences and 0 otherwise.

Our measure for efficiency is then given by the (expected) increase in payoff over the minimum possible payoff that the players achieve, relative to the maximum increase they could possibly achieve. We call this measure the expected efficiency,

$$E[E_p] = \frac{E[S_p^{\text{real}}] - S_p^{\min}}{S_p^{\max} - S_p^{\min}}.$$

This measure is then unaffected by the outcome of the random draw which takes place if both players state equal preferences. For a single project, the denominator will be zero if both players have the same true preference, so the expected efficiency would not be properly defined. We will, however, only consider aggregate measures (across periods), such that this problem does not occur in practice. Expected efficiency will be computed based on aggregates, i.e. we first sum over the maximum, minimum, and expected realized payoffs and then calculate the expected efficiency as follows (where $1, \ldots, P$ denote the projects over which we aggregate),

$$E[E] = \frac{\sum_{p=1}^{P}(E[S_p^{\text{real}}] - S_p^{\min})}{\sum_{p=1}^{P}(S_p^{\max} - S_p^{\min})}.$$

Note that if players always state their preferences truthfully, then $E[E] = 1$ and if they follow the stage–game Nash–equilibrium strategy to always state a strong preference, then $E[E] = \frac{1}{2}$.

3 Theoretical Predictions and Hypotheses

3.1 The Benefits from Cooperation

We start by illustrating the possible benefits from linking decisions (either through exogenous budgets or through conditionally cooperative behavior). We present the possible efficiency gains from honest behavior as compared to (stage–game Nash–)equilibrium play. Suppose (analogously to the choice of parameters in our experimental setting) that the two preference intensities satisfy $s = 2w$. Denote by $EU(x, y)$ an agent's expected payoff from behavior x if the other plays y, $x, y \in \{h, s, w\}$, where h stands for honest behavior, and s (w) for always reporting strong (weak) preferences.[13] The expected payoffs are displayed in Table 4, for the computations see the appendix.

	h	s	w
h	$\left(\frac{7}{8}w, \frac{7}{8}w\right)$	$\left(\frac{1}{2}w, \frac{9}{8}w\right)$	$\left(\frac{5}{4}w, \frac{3}{8}w\right)$
s	$\left(\frac{9}{8}w, \frac{1}{2}w\right)$	$\left(\frac{3}{4}w, \frac{3}{4}w\right)$	$\left(\frac{3}{2}w, 0\right)$
w	$\left(\frac{3}{8}w, \frac{5}{4}w\right)$	$\left(0, \frac{3}{2}w\right)$	$\left(\frac{3}{4}w, \frac{3}{4}w\right)$

Table 4: Expected payoffs from being honest, tough, and nice, first number for row player, second for column player.

As it turns out, honest play by both agents increases the expected total payoff by approximately 16.7 %, relative to the (stage–game) equilibrium payoff. However, if the other player is always reporting honestly, the incentive for one player to always state a strong preference is higher than the efficiency gain from mutual honest behavior (i.e. it raises his payoff by 28.6 % compared to being honest as well). Finally, observe that given the other player always reports a strong preference, doing the same even increases a player's payoff by 50 % compared to honest behavior.

3.2 Equilibria

As argued above, in the one–shot game it is a dominant strategy to always report a strong preference. This is also the prediction for each repetition of the stage game if matching is random and players cannot identify players they interact(ed) with. Thus the only equilibrium in RAN is that players always state a strong preference.

[13] We are slightly abusing the notation here, by using s both to denote a strong preference and the strategy to state a strong preference irrespective of the true preference (and correspondingly for w).

For treatment EXO, it follows from Jackson and Sonnenschein (2007) that it is a Bayesian equilibrium to state preferences as truthfully as possible within the budget. That is, a player would state his true preference until either the remaining budget for strong preferences is 0 (in which case he has to state a weak preference in all remaining periods regardless of his true preference) or equals the number of remaining periods (in which case he can state a strong preference for all remaining periods). Unless players draw extreme sequences of preferences, in this equilibrium preferences are generally reported truthfully except possibly the last few periods. Honesty and expected efficiency would hence almost be equal to those under truthful representation, i.e. $E[E] \approx 1$.

In treatments FIX and RLK, we observe that in the last period, there is no reason to deviate from the stage–game dominant strategy to report a strong preference. By a standard backward–induction argument, the same holds for the finitely repeated game. Hence the only subgame-perfect Nash equilibrium is to report a strong preference for each project independent of the true preference.

It is, however, a well–established result that experimental subjects do generally not perform backward induction over many stages. Thus, they may approach a game that is repeated many times like an infinitely repeated game until end–game effects kick in close to the end. It is therefore reasonable to consider equilibria for infinitely repeated games. Following from Folk–theorem arguments, there are many possible equilibria that players could follow and this gives rise to a selection problem. For treatment FIX, we show below that there is a relatively simple equilibrium where players manage to link the decisions in two periods. For treatment RLK, we find a similar equilibrium where players link the projects in each period and take information about all the players in the group into account. For CMP we will see that even in the finitely repeated game there is an equilibrium that exploits the opportunity to link projects.

We start by considering FIX. While there are many more equilibria, the one suggested here is the most simple one that makes use of the repeated game to achieve some efficiency gains through linking decisions. Given that coordinating on any such equilibrium is difficult, this arguably most simple equilibrium that allows for limited efficiency gains seems an appropriate benchmark to assess how well experimental subjects exploit the opportunities to endogenously link decisions. In contrast, it would clearly not come as a surprise if players do not manage to coordinate on highly complicated nearly efficient equilibria. For simplicity, we consider payoffs to be $w = 1$ and $s = 2$ in the following.

PROPOSITION 1 (FIX) *In the infinitely repeated game with fixed matching between two players (as in FIX), the following constitutes a subgame–perfect Nash equilibrium.*[14] *Players consider periods in (disjoint) blocks of two and adopt the following trigger strategy: state one strong and one weak preference (truthfully if possible, randomly otherwise) in a block of two periods if both players have done so in the past. If one player deviates, always state strong preferences for 16 periods, then go back to the pattern of one strong and one weak preference. The ex–ante (i.e. before preferences are drawn) expected payoff per period in this equilibrium is $\frac{13}{16}$. The ex–ante expected honesty rates are $H_w = H_s = \frac{3}{4}$.*

PROOF See Appendix B. □

The ex–ante expected payoff when both players are always honest (which leads to the maximum possible payoff) is $\frac{7}{8}$, while the ex-ante expected payoff in the stage–game Nash equilibrium is $\frac{3}{4}$ (see Appendix A). Thus the above equilibrium would allow the players to obtain half of the possible efficiency gains of perfect honesty compared to the stage–game Nash equilibrium, or $E[E] = \frac{3}{4}$.[15]

For treatment RLK the situation is more difficult, as the number of projects per period for an individual player is random and in particular, the number of projects per period can be uneven. On the other hand, the fact that players frequently have more than one project per period makes it possible to link these projects and allows for some efficiency gains from linking without having to link projects across periods. We thus consider such

[14]We ignore discounting for the moment. In an experiment, it is plausible to assume that players do not discount later periods, because they obtain the payoffs for all periods only at the end of the experiment. Thus, it makes sense to consider players treating the game as if it had an infinite horizon, but without discounting, even though in a theoretical model this implies conceptual problems of comparing infinite payoffs. Here it has the advantage to give us a minimal length of a punishment phase. We will address discounting below.

[15]For any specific set of preferences, $E[E]$ might be higher or lower in the given equilibrium, but in expectation it equals $\frac{3}{4}$.

an (infinitely repeated game) equilibrium that makes use of the opportunity to link several projects within one period, but would not link projects across periods. Again, the reason to consider this equilibrium is that it is arguably the easiest and the efficiency gains are relatively small, so that it is not too demanding a benchmark to compare the experimental behavior to.

A crucial aspect of RLK is that any deviation from a cooperative strategy can be observed by all three other players in the group and hence can attract relatively harsh punishment. Furthermore, punishment can be targeted to the deviator and thus cooperation through honesty among other group members does not have to be ended as a result of a deviation by one player even if they play a grim–trigger strategy. To ease the derivation of the equilibrium, we will, however, consider a trigger strategy that leads to temporary complete breakdown of cooperation. The incentives for a player not to deviate would be the same and hence the length of the punishment phase would not be affected if punishment was targeted to him, but it would complicate the analysis for the other players on the punishment path.

Given that in expectation a player has a strong preference for half of his projects, a reasonable assumption is that to avoid punishment a player with two projects in a period or with four projects would be required to state half of the preferences as weak. In case the number of projects is one or three, we assume the milder of two obvious possible requirements, namely that for one project there is no restriction, and for three projects, the player has to state only one weak preference. Requiring to state one and two weak preferences, respectively, would not lead to additional efficiency gains, but would make deviation more likely.

Proposition 2 (RLK) *In the infinitely repeated game with random link formation in a fixed group of four players (as in RLK), the following constitutes a subgame–perfect Nash equilibrium.*[16] *Players consider one period at a time and adopt the following trigger strategy: for one project, state a strong preference, for two or three projects, state one weak preference and for four projects, state two weak preferences if all players have followed this pattern in*

[16]We again ignore discounting for the moment.

the past. State preferences as truthfully as possible given these restrictions. If one player deviates, always state strong preferences for all projects for 19 periods, then go back to the above pattern. The ex–ante expected payoff per period in this equilibrium is $\frac{695}{432}$. The ex–ante expected honesty rates are $H_w = \frac{127}{216} \approx 58.8\%$ and $H_s = \frac{61}{72} \approx 84.7\%$.[17]

Proof See Appendix B. □

If players state preferences honestly, the expected payoff per project is $\frac{7}{8}$ (see Appendix A). Since each player has in expectation two projects, the expected payoff per period given truthful representation is thus $\frac{7}{4}$. The ex–ante expected payoff in the stage–game Nash equilibrium is $\frac{3}{4}$ per project, so $\frac{3}{2}$ per period. Thus the above equilibrium would allow the players to obtain $\frac{695-648}{756-648} = \frac{47}{108} \approx 43.5\%$ of the possible efficiency gains of perfect honesty compared to the stage–game Nash equilibrium.

Remark 1 *If we introduce discounting, we see that for a sufficiently high discount factor, a grim–trigger strategy that switches to always stating a strong preference but is otherwise as above, yields a subgame–perfect Nash equilibrium in both FIX and RLK. For simplicity, assume for FIX that there is no discounting between the two periods in a block and let δ denote the discount factor between two blocks. Then deviating does not pay if the discounted loss of entering the punishment phase is larger than the (maximum) immediate gain from deviating, or $\sum_{i=1}^{\infty} \delta^i \frac{1}{8} > 1 \Leftrightarrow \frac{\delta}{1-\delta} > 8 \Leftrightarrow \delta > \frac{8}{9}$. For RLK, since only projects in one period are linked, we can consider discounting in the standard way, so δ denotes the discount factor between two periods. Then deviating does not pay if the discounted loss of entering the punishment phase is larger than the (maximum) immediate gain from deviating, or $\sum_{i=1}^{\infty} \delta^i \frac{47}{432} > 2 \Leftrightarrow \frac{\delta}{1-\delta} > \frac{864}{47} \Leftrightarrow \delta > \frac{864}{911}$.*

Remark 2 *We could observe similar equilibria for FIX and RLK even for the finitely repeated game if there are (or players assume that there are) some intrinsically honest*

[17] $H_w < H_s$ in this equilibrium, because players who have one or three projects state more strong than weak preferences, so strong preferences get more honestly represented than weak ones.

players[18]*, who would, however, punish dishonest others. In these equilibria, players would try to signal honesty by not excessively overstating preferences. Again, many such equilibria are possible, depending on what would be considered unmistakable evidence that a player is dishonest. The most simple equilibrium would again involve linking the decisions in a block of two periods in FIX and linking the decisions in one period in RLK. Honesty would decline towards the end and this would happen earlier the lower the assumed share of intrinsically honest players.*

The theoretical analysis of treatment CMP differs in a crucial aspect from that of FIX and RLK. As we will see, the choice of partners allows for the linking of decisions to emerge in a subgame–perfect equilibrium even in the finitely repeated game.[19] Thus in CMP the prediction that players might be able to endogenize a budget does not rely on the assumption that they treat the game as if it were infinitely repeated (or there were some intrinsically honest players that others try to mimic). We again consider only minimal linking, that is the linking of two projects within a period rather than more efficient linking across several periods. The motivation is again that this is the least demanding equilibrium that makes use of the opportunity to link projects and thus establishes an arguably appropriate benchmark for the experimental behavior.

Proposition 3 (CMP) *In the finitely repeated two-stage game where players first choose a partner for the project originating with them and then state preferences for all their projects (as in CMP), the following constitutes a subgame–perfect Nash equilibrium.*[20] *Let the number of periods be n. Players consider one period at a time and adopt the following trigger strategy: in the first stage, two pairs form endogenously, that is two players, respectively, choose each other and hence share two projects.*[21] *In the second stage, both players*

[18]For example, this could be players who actually care for efficiency as suggested by Charness and Rabin (2002) and Engelmann and Strobel (2004).

[19]Ule (2006) presents a more general analysis for prisoner's dilemmas with corresponding results.

[20]Since we consider the finitely repeated game, we can ignore discounting.

[21]There is obviously a coordination problem in practise, but over the course of a few periods, this should

in each pair state one strong and one weak preference. Both players continue in this fashion until period $n-2$ if they have done so in the past. In period $n-1$, they will state two strong preferences if they have indeed two strong preferences but again one weak and one strong otherwise. In period n, if no player stated two strong preferences before, they still choose each other but state two strong preferences. If player i deviates by stating preferences other than one strong and one weak in any period, then for the rest of the game, player j (the "partner" of i) chooses another player, while player i continues to choose j and they both state strong preferences.[22] *If a player is chosen by a player from a different pair, he states a strong preference for the resulting project. The ex–ante expected payoff per period in this equilibrium is $\frac{13}{8}$. The ex–ante expected honesty rates are $H_w = H_s = \frac{3}{4}$.*

PROOF See Appendix B. □

As in the equilibrium for the infinitely repeated game of FIX, in this equilibrium, always two projects are linked and thus the ex–ante expected payoff per period (except for the last two periods) is $\frac{13}{8}$. The expected payoff from two projets and perfect honesty is $\frac{7}{4}$ and in the stage–game Nash equilibrium it is $\frac{3}{2}$. Hence as for FIX, the above equilibrium allows players to reap half of the efficiency gains that would be obtained from perfect honesty compared to the stage–game Nash equilibrium.

Even though calculating the subgame–perfect Nash equilibrium is certainly demanding for the subjects, its logic is fairly obvious. In addition to keeping an endogenously formed pair intact, players might also want to appear honest because in the initial phase where pairs are formed, one would want to attract a potential partner and furthermore, they might attract even an additional project if the remaining two players do not form an equilibrium pair. Thus in CMP, apparent honesty not only possibly triggers reciprocal honesty but can also attract partners and hence increase the number of projects, which directly increases the expected payoffs.

---

be relatively easy to sort out.

[22]There are obvious alternatives for the punishment path. In particular, both players could choose different partners. We focus on this pattern since it minimizes the number of cases to be considered.

To summarize, we have an almost efficient, almost perfectly honest subgame–perfect equilibrium in EXO. For CMP, we have a relatively simple subgame–perfect equilibrium with endogenous pair formation and limited honesty and efficiency gains. For FIX and RLK we have such equilibria only for the infinitely repeated game (or as signalling equilibria), hence these are applicable only if players perceive the game as infinite (or if there are some truly honest players). Hence it appears more likely that players will approximate the behavior of an equilibrium with partial linking in CMP than in RLK and FIX. For RAN, repeated dominant–strategy play is the only plausible theoretical prediction.

3.3 Hypotheses

Based on our theoretical results, we derive the following hypotheses.

(H1) With random matching (RAN) subjects are expected to play the stage–game equilibrium, therefore

$$H_w(RAN) \approx 0,\ H_s(RAN) \approx 1,\ E[E] \approx \frac{1}{2}.$$

(H2) With exogenous budgets (EXO) subjects are expected to use their budgets rationally and therefore should approximately reach honest representation and efficiency.

$$H_w(EXO) \approx 1,\ H_s(EXO) \approx 1,\ E[E] \approx 1.$$

(H3) Fixed matching in pairs (FIX) or groups of four (RLK) should lead to more honest representation of weak preferences than random matching (RAN), but to less honest representation than exogenous budgets (EXO). Efficiency in FIX and RLK should thus be between the levels in RAN and in EXO.

$$\begin{aligned} &H_w(RAN) < H_w(FIX) < H_w(EXO)\\ &H_w(RAN) < H_w(RLK) < H_w(EXO)\\ &E[E](RAN) < E[E](FIX) < E[E](EXO)\\ &E[E](RAN) < E[E](RLK) < E[E](EXO). \end{aligned}$$

(H4) Competition for partners should lead to more honest representation of weak preferences than RLK and FIX, but to less honest representation than in EXO. Representation of strong preferences should be less honest than in FIX and RLK. Efficiency in CMP should be higher than in FIX and RLK, but lower than in EXO.

$$H_w(RLK), H_w(FIX) < H_w(CMP) < H_w(EXO)$$
$$H_s(CMP) < H_s(RLK), H_s(FIX)$$
$$E[E](RLK), E[E](FIX) < E[E](CMP) < E[E](EXO)$$

4 Results

In the following we first compare honesty rates in the different experimental treatments. We will then examine how this translates into differences in efficiency.

4.1 Results - Honesty

The average levels of truthful representation of strong and weak preferences are found in Tables 5 and 6.

	H_w		H_s	
treatment	data	EQ prediction	data	EQ prediction
RAN	7.0%	0%	97.2%	100%
EXO	85.7% (88.7%)	100%	84.9% (89.3%)	100%

Table 5: Share of truthfully represented weak and strong preferences for RAN and EXO, data and ex–ante equilibrium predictions. The numbers in parentheses for EXO correct for forced lies due to a depleted budget (H_s) or for free lies due to a full budget (H_w).

Let us first look at the behavior in treatments RAN and EXO (Table 5). As expected, players in RAN generally state a strong preference, i.e. they misrepresent their preferences if they are weak (truthful representation in only 7% of the cases where a weak preference was observed) but truthfully represent their preferences when they are strong. Indeed 12 out of 24 subjects always state a strong preferences irrespective of their true preference, another five always state a strong preference when this is their true preference. Hence, we find clear support for our first hypothesis and establish that the incentive problem is empirically relevant. While this result should not come as a surprise to a theorist, it is noteworthy that the rates of truthful representation of weak preferences (which can be seen as a measure of cooperativeness) is lower than cooperation rates that have been observed, for example, in prisoner's dilemma games with random matching.[23]

In the treatment with exogenous budgets, the picture is remarkably different. Players overwhelmingly report their preferences truthfully, supporting our second hypothesis. It is interesting to note that the rate of truthful representation of strong preferences is substantially lower than in RAN (84.9%). Partly this is due to the fact that the budget becomes binding for some players in the last few periods, i.e. they are forced to "lie downwards". Even if we correct for this, however, the share of truthfully represented strong preferences rises to only 89.3%. Statements in the post–experimental questionnaire indicate that some subjects became worried of spending their budget too fast when they had many strong preferences in the first periods and wanted to save their budget for later. Similarly, if we correct for "free lies", i.e. when players have a sufficient budget to state a strong preference in all the remaining periods, H_w increases from 85.7% to 88.7%. Overall we find that the Jackson–Sonnenschein mechanism works remarkably well as it achieves almost perfectly truthful revelation (although, even after correcting for forced and free lies, only 4 out of 24 players *always* report their preferences truthfully). According to a Mann–Whitney test, using the matching groups as independent observations, aggregating across all pairs and all periods within each matching group, in EXO H_w is significantly higher and H_s is significantly lower than in RAN ($p = 10\%$, note that for three independent observations per treatment, this is the smallest possible p.)

We summarize the main results from the baseline treatments in Results 1 and 2.

RESULT 1 *The incentive problem is empirically relevant since in RAN subjects overwhelmingly play their dominant strategy to state a strong preference.*

[23] Cooper et al. (1996) report cooperation rates of 22% in the last ten periods of a prisoner's dilemma game with random matching across 20 periods.

Result 2 *The Jackson–Sonnenschein mechanism (EXO) achieves a significant improvement. Stated preferences are overwhelmingly truthful. Subjects partly understate their preferences, either because a depleted budget forces them to do so or because they are afraid of spending it too quickly.*

We now turn to treatments FIX, RLK and CMP (see Table 6), to investigate how well social interaction can help to overcome the incentive constraints without an exogenously enforced budget.

	H_w		H_s	
treatment	data	EQ prediction	data	EQ prediction
RAN	7.0%	0%	97.2%	100%
EXO	85.7% (88.7%)	100%	84.9% (89.3%)	100%
FIX	11.6%	75%	97.9%	75%
RLK	12.8%	58.8 %	98.2%	84.7%
CMP	30.9%	75%	93.8%	75%

Table 6: Share of truthfully represented weak and strong preferences for FIX, RLK, and CMP, data and ex–ante equilibrium predictions.

The honesty rates in FIX and RLK do not differ substantially or significantly from those in RAN. Essentially, it appears that repeated interaction has hardly any effect on the honesty of represented preferences. Thus we do not find support for our third hypothesis.

In contrast, competition has a notable effect on the honesty rates. Weak preferences are represented honestly in 30.9%, in contrast to 12.8% in the random links treatment. H_w is significantly higher in CMP than in RLK, FIX, or RAN (Mann–Whitney, $p < 5\%$, using aggregate measures for groups of four, fixed pairs, or matching groups as independent observations). Furthermore, strong preferences are represented significantly less honestly in CMP than in FIX or RLK (Mann–Whitney, $p < 5\%$) or RAN ($p < 10\%$). That means, in line with our fourth hypothesis, in CMP players state a weak preference more frequently than in RAN, FIX, and RLK, both if their true preference is weak and if it is strong.[24]

[24]How the players actually choose their partners in CMP will be discussed below.

Although these differences are significant, they are substantially smaller than the effect of the exogenous budget. In particular, H_w is significantly smaller and H_s significantly larger in CMP than in EXO (Mann–Whitney, $p < 5\%$), as predicted.

If we consider the honesty rates relative to the equilibrium predictions, we again find that H_w is substantially and significantly higher in CMP (41.2%), than in RLK (21.8%), ($p < 10\%$), and FIX (15.5%), ($p < 5\%$), but less than half that of EXO, ($p < 5\%$). Strong preferences are stated more honestly relative to the equilibrium to a similar degree in CMP (125%), in RLK (115.9%) and in FIX (130.5%) (the differences are actually all significant at $p < 5\%$).

We summarize these observations as follows.

Result 3 *Repeated interaction in pairs (treatment FIX) or in groups of four without the chance of choosing partners (treatment RLK) has essentially no effect on honesty rates.*

Result 4 *If players can choose partners for the interaction (treatment CMP), this significantly increases the rate of truthfully stated weak preferences and significantly reduces the rate of truthfully stated strong preferences compared to treatments RAN, FIX, and RLK. These effects, are however, substantially and significantly weaker than those induced by the exogenous budgets (EXO).*

To summarize, the types of social interaction that we investigate are by far not as effective as the Jackson–Sonnenschein budgeting mechanism in encouraging honest statements of preferences.

4.2 Results - Efficiency

In this section, we report how the different behavior translates into efficiency differences among the treatments. The expected efficiency aggregated across all periods is presented in Table 7 alongside the aggregate across the first ten periods.

As we can see, most treatments achieve efficiency levels only slightly above the stage–game Nash–equilibrium value of 50%, whereas the exogenous budget treatment almost reaches full efficiency. Correcting for the fact that depleted budgets force players to state

treatment	Expected Efficiency	Exp. Eff. Periods 1-10	Equil. Exp. Eff.
RAN	52.4%	58.4%	50%
EXO	87.2% (89.1%)	91.2%	100%
FIX	55.0%	55.8%	75%
RLK	55.9%	59.4%	71.8%
CMP	62.1%	70.3%	75%

Table 7: Expected efficiency in the different experimental treatments. The number in parentheses for EXO corrects for forced or free lies due to a depleted or full budget. The third column shows the expected efficiency in Periods 1 to 10 and the fourth column the ex–ante expected efficiency in equilibrium.

weak preferences when their true preferences are strong, about 80% of the possible efficiency gains compared to the stage–game Nash equilibrium are achieved. The expected efficiency is significantly higher than in each of the other treatments (Mann–Whitney tests, $p \leq 10\%$). Among the other treatments, only the competition treatment achieves significantly higher expected efficiency than treatments RAN, FIX, and RLK (Mann–Whitney tests, $p < 10\%$). In no other pair of treatments does expected efficiency differ significantly. Compared to the difference in the rates of truthful representations H_s and H_w, the differences in expected efficiency between CMP and the other treatments are relatively small. This is so because the efficiency gains due to more truthful revelation of weak preferences are partly compensated by the more frequent misrepresentation of strong preferences. In line with our fourth hypothesis, however, we find that the positive effect dominates.

In all treatments, there are initially more attempts to cooperate (or also possibly more errors) such that the expected efficiency in the first ten periods is higher than in the overall data. As shown in Table 7, the effect is, however, smaller in FIX and RLK than in RAN and largest in CMP. As a result, in periods 1 to 10, the difference between expected efficiency in CMP and in RAN, FIX, and RLK is now significant even at $p < 5\%$ (Mann–Whitney). Indeed, in all treatments, H_w is higher, but in RAN, FIX and RLK by only 4 to 6 percentage points, while it reaches 43.5% in CMP (compared to 30.9% across all periods.) Hence the differences between treatments are initially stronger.[25]

If we consider the efficiency gains relative to the gains in the suggested benchmark equilibria compared to the stage–game Nash equilibrium, we find that in EXO 74.4% of the equilibrium efficiency gains are captured, while it is only 20% in FIX and 27.1% in RLK, but 48.4% in CMP. CMP differs significantly neither from RLK nor EXO ($p > 10\%$), while CMP still differs significantly from FIX ($p < 10\%$), and EXO from RLK ($p < 10\%$) and FIX ($p < 5\%$). Thus even relative to the more modest plausible equilibrium benchmarks, only CMP is reaching substantial efficiency gains, while these are still much lower than in EXO.

There is also a lot of heterogeneity among groups within treatments. For example, one group in RLK achieved 81.3% of the equilibrium efficiency gains, while no other group reaches more than 40%. In FIX, one pair reaches 71.4% of equilibrium efficiency gains, but three pairs obtain exactly the stage–game Nash efficiency. CMP shows generally higher efficiency levels with three of seven groups reaching more than 60% of equilibrium efficiency gains and even six being above 25%. In contrast, in EXO all three independent groups obtain 72% or more of the equilibrium efficiency gains (if we control for depleted budgets even more than 76%.)

Summarizing our results, we find clear support for our hypotheses that random matching without exogenous budgets leads to nearly stage–game Nash–equilibrium play (Hypothesis 1) and that exogenous budgets are most effective in increasing truthful representation of weak preferences and that this leads to nearly full efficiency (Hypothesis 2). We also find support for the hypothesis that weak preferences are stated more truthfully and strong preferences less truthfully in CMP than in RLK and FIX and that this translates into higher expected efficiency (Hypothesis 4). Contrary to Hypothesis 3, fixed matching in pairs or groups of four and the possibility of reputation building has a significant effect neither on honesty rates nor on expected efficiency. Based on these (partly surprising) results, in the next section we try to isolate what drives successful cooperative behavior in our environment of two–sided private information.

[25] In all treatments, H_s is marginally smaller in the first ten periods and in EXO, H_w is marginally higher.

5 What Drives Successful Cooperation?

Our above results show that in our experimental setting social interaction has little or no effects on the representation of preferences unless players can choose their partners. At a first glance, this appears to be surprising, since in simpler games (like trust games or prisoner's dilemma games) repeated interaction, like in FIX, usually increases cooperation substantially. In other experiments, in settings similar to RLK, the opportunity to build a reputation enables subjects to cooperate (for example in trust games or helping games, see, e.g., Engelmann and Fischbacher, 2003). While we find that the choice of partners has a significant effect on efficiency, they are not as dramatic as, e.g., in the trust game (see Huck, Lünser, and Tyran, 2006).

We will now address possible explanations for the observed failure of cooperation in FIX and RLK and the relatively low efficiency gains in CMP. Two further treatments will help us evaluate these explanations. The most plausible reasons are the following:

1. The signals are ambiguous. A player cannot observe whether the other player lied or not and does not know whether repeatedly stated strong preferences are a true reflection of randomly chosen preferences or the result of exaggeration. As a result, no simple strategy like tit–for–tat in honesty is possible. A conditionally cooperative strategy can only use stochastic information. This implies the next problem.

2. Players clearly face a coordination problem. Even if they want to play conditionally cooperative, they have to implicitly agree which horizon is chosen to judge the other's honesty. That means one player has to know in what cases the other will judge his behavior as a sign of dishonesty and will revert to punishment. Learning the other's strategy (or the others' strategies in RLK and CMP) would take a considerable number of periods for experimentation. Put differently, while a conditionally cooperative strategy implies a budget for the other player, it is far more difficult to communicate than an exogenous budget.

3. Finally, understanding the possible gains from cooperation and understanding that it is possible to play a conditionally cooperative strategy based on stochastic information is intellectually relatively demanding and both players (or all four in RLK) must understand this in order to coordinate on cooperation through mutual honesty.

Note that the last problem is substantially reduced if players can choose each other as in CMP. Here, if two out of four players understand how to reap gains from cooperation, they can choose each other and endogenously form a cooperating pair. We will discuss below that this is exactly what happens in several groups and that this drives the observed differences in behavior. First, we describe two control treatments that we ran in order to test explanations number 1 and 2 above.[26]

Control Treatment I: Fixed pairs with ex–post complete information, FXI The treatment is identical to "Stable Partnerships" (FIX), except for the fact that, at the end of each period, in addition to the stated preferences, both players observe also the true preference of the other player. This implies that honesty is now directly observable, so reciprocity does not have to rely on stochastic methods and there is no need to budget decisions.

Control Treatment II: Fixed pairs with multiple projects per period, F4P The treatment is identical to "Stable Partnerships" (FIX), except for the fact that each pair decided on four independent projects each period. This could enable the players to overcome the problem of coordinating on a specific conditionally cooperative strategy and hence on the number of projects that should be linked as it suggests to link the four projects within each period. The payoffs for strong (weak) preferences where reduced to 20p (10p) in order to compensate for the higher number of projects and leave hourly wages comparable to the other treatments.

The first problem discussed above is in principle eliminated if players can ex–post observe each others' true preferences. In this case, they can clearly assess each others' honesty and hence have sufficient information to play simple strategies like tit–for–tat in honesty. Our control treatment FXI implements such ex–post information. Indeed if we again consider that players perceive the game as infinitely repeated, we can easily establish honest representation as an equilibrium.

[26]We ran one session with 14 subjects of each treatment. Given the fixed matching this yields 7 independent observations each.

Proposition 4 *In the infinitely repeated game with fixed matching between two players and ex–post information about the intensity of preferences (as in FXI), the following constitutes a subgame–perfect Nash equilibrium.*[27] *Players state their preferences truthfully if both players have done so in the past. If it becomes known at the end of a period that a player misrepresented his preference, both players state always strong preferences for four periods and then go back to honest representation. The ex–ante expected payoff per period in this equilibrium is $\frac{7}{8}$, and the ex–ante expected honesty rates are $H_w = H_s = 1$.*

Proof See Appendix B. □

Concerning the second problem discussed above, coordinating on a budget should be facilitated in our second control treatment F4P where, within a fixed pair, players decide about four independent projects simultaneously in every period. This suggests to assign each other a budget per period, most likely either two or possibly at most three strong stated preferences. Deviations from this budget could then be retaliated in the next period(s). Such a strategy could relatively easily be signaled and would lead already to substantial improvements in truthful revelation of weak preferences (but would also imply a decrease in the truthful revelation of strong preferences). While it would further increase truthful revelation if projects were linked across periods, this would again be difficult and hence we would expect short–term budgeting. We consider here the case of linking only the projects within a period and assigning a budget of two strong preferences.

Proposition 5 *In the infinitely repeated game with fixed matching between two players and four projects in each period (as in F4P), the following constitutes a subgame–perfect Nash equilibrium.*[28] *Players play the following trigger strategy: state two strong and two weak strategies (as truthfully as possible, randomly otherwise) in each period if both players have done so in the past. If one player deviates, always state strong preferences for seven periods, then go back to the pattern of two strong and two weak preferences. The ex–ante*

[27] We again ignore discounting for the moment.

[28] We ignore again discounting for the moment.

expected payoff per period in this equilibrium is $\frac{53}{16}$*. The ex–ante expected honesty rates are* $H_w = H_s = \frac{13}{16}$.

Proof See Appendix B. $\square$

Note that if the players perceive the game as infinitely repeated in FXI, they reach the maximum expected efficiency in a relatively straightforward equilibrium. In F4P the expected payoff of perfect honesty is $4 \cdot \frac{7}{8} = \frac{7}{2}$, while the expected payoff in the stage-game Nash equilibrium is $4 \cdot \frac{3}{4} = 3$, so the expected efficiency gain in the above equilibrium is $(\frac{53}{16} - 3)/(\frac{7}{2} - 3) = \frac{5}{8}$ of the gain that would be reached through perfect honesty compared to stage–game Nash behavior. The expected efficiency gain is larger than in the discussed equilibria of CMP or FIX, because four projects are linked and not just two.

If we again consider discounting, it is straightforward to see that we can establish the path in FXI as equilibrium through a grim–trigger strategy if $\delta > \frac{4}{5}$ and in F4P if $\sum_{i=1}^{\infty} \delta^i \frac{5}{16} > 2 \Leftrightarrow \frac{\delta}{1-\delta} > \frac{32}{5} \Leftrightarrow \delta > \frac{32}{37}$.

treatment	H_w	H_s	Exp. Efficiency
RAN	7.0% [0%]	97.2% [100%]	52.4% [50%]
EXO	85.7% (88.7%) [100%]	84.9% (89.3%) [100%]	87.2% (89.1%) [100%]
FIX	11.6% [75%]	97.9% [75%]	55.0% [75%]
RLK	12.8% [58.8%]	98.2% [84.7%]	55.9% [71.8%]
CMP	30.9% [75%]	93.8% [75%]	62.1% [75%]
FXI	5.6% [100%]	98.8% [100%]	52.4 [100%]
F4P	19.1% [81.3%]	92.4% [81.3%]	56.6 [81.3%]

Table 8: Share of truthfully represented weak and strong preferences and expected efficiency for FXI and F4P. In brackets the ex–ante predicted equilibrium values for each treatment.

Table 8 relates the honesty rates and expected efficiency in the two control treatments to those in the other treatments and the equilibrium predictions. There is essentially no effect of the ex–post information in FXI. If anything, it leads to even more frequently stated strong preferences. As a result, efficiency is virtually the same as in FIX. Indeed, pairs are locked even quicker in a "strong–strong" state than in the other treatments. It appears that some players actually followed a grim–trigger strategy, reverting to constantly stating strong preferences after they observed the other "lied upwards" only once. The failure to achieve cooperation in this treatment seems to be driven again by coordination problems and by slower understanding of some players. The subjects in our experiment generally seem to be impatient and unforgiving. In most treatments, according to post-experimental questionnaires, some subjects overestimate their own honesty and underestimate the honesty of others. They then punish people for being dishonest that are not less honest than they are themselves. Interestingly, this effect does not even disappear if subjects can actually judge the other's honesty equally well as their own.

As we can also see from Table 8, in F4P there is some increase in H_w compared to RAN or FIX, but the effect is weaker than in CMP. Indeed, only the difference with FXI is significant ($p < 5\%$). There is also a notable decrease in H_s. Since the increase of H_w is partly canceled by the decrease in H_s, the increase in efficiency compared to FIX is very small and insignificant.[29] In particular, if we consider again the efficiency gains relative to the equilibrium gains compared to the stage–game Nash equilibrium, we find that only 21.1% of these gains are captured, virtually identical to the result for FIX.

Therefore, it appears that substantially reducing the coordination problem alone does not enable pairs of players to coordinate on relatively truthful representation of preferences. Even when measured against the more reasonable benchmark of the equilibrium where in each period subjects are restricted to a 2–strong–2–weak–budget, our subjects exaggerate their preferences quite frequently, but state strong preferences more honestly.[30]

So why do subjects on average represent weak preferences more truthfully in the competition treatment? The main reason appears to be that the competition treatment reduces

[29] If we again restrict attention to the first ten periods, we find expected efficiency only marginally increased compared to the complete data, to 52.6% in FXI and to 59.7% in F4P. This is driven by a larger H_w (29.7% in F4P, 12.3% in FXI), but partly compensated by H_s being marginally smaller in both treatments.

[30] We also calculated the honesty rates given the draw of preferences observed in the experiment and the equilibrium strategy. This would yield honesty rates of $H_w = 81.0\%$ and $H_s = 79.8\%$, very close to the ex–ante expected rates.

the third, and partly also the second of the problems that we stated at the beginning of this section. If two out of four players understand the gains from mutually truthful representation of preferences, they can signal this by stating some weak preferences. They can then choose each other as partners for their projects. Hence the third problem is reduced. If two subjects in a group of four see the way to reap gains from cooperation, this will at least lead them to truthful representation. We see indeed very clear examples of this kind of endogenous pairing in three of our seven groups. Interesting is the reaction of the remaining two players. Partly, they also choose these two players, because the latter state a weak preference more frequently, allowing the former to gain a higher payoff. In one group, however, the remaining two players choose each other, but always state a strong preference.[31]

The competition treatment also partly solves the second problem. If a pair forms endogenously, they share two projects each period. Thus, as suggested by the subgame–perfect equilibrium presented above, they can link these two projects and hence allow each other to state only one strong preference per period.

Another cooperation facilitating property of this treatment is that it allows unambiguous punishment. In the other treatments, the only way to punish a player is to state a strong preference in the next interaction with him. This, however, is not clearly seen as punishment since it could also just be the truthful representation of a strong preference. In CMP, however, once a pair has formed endogenously, one player can punish her partner by choosing another partner for a limited time.[32] Moreover, punishment is a credible threat in the subgame–perfect equilibrium.

We have seen that the only treatment without exogenous budgets that achieves substantial levels of honest representation of weak preferences is CMP and we had argued that this is driven by the possibility to attract more partners or to form endogenous pairs that

[31]One of these states in the questionnaire, that he considered it unfair to be left out by these two players and hence started choosing the remaining player, apparently missing the reason why the other two chose each other.

[32]One subject stated in the questionnaire, that she followed the strategy to state one–weak–one–strong and if her partner deviated from this rule, she would switch to another partner for one period. The data shows that she indeed did.

achieve higher efficiency through truthful representation. An important question is whether this strategy pays off. There is some evidence suggesting that it does. If we compare group by group the average total payoff of the two subjects with the highest H_w to that of the two subjects with the lowest H_w, we see that the former is higher than the latter in five of the seven groups and according to a Wilcoxon signed-rank test, this difference is significant ($z = 1.693, p < 10\%$).[33] Furthermore, we observe three groups where consistently two players choose each other and are relatively honest (H_w between 42% and 77%). These three groups are also those with the highest expected efficiency and indeed they manage to reap between 63% and 74% of the efficiency gains that would be obtained in the cooperative subgame–perfect equilibrium compared to the stage–game Nash equilibrium. This compares to between 11% and 48% in the other groups. Therefore, building endogenous pairs paid off.

6 Conclusion

We have investigated the behavior of experimental subjects in a simple voting game with private information about the intensity of preferences. We have seen that the exogenously enforced budgeting mechanism as suggested by Jackson and Sonnenschein (2007) works very well in inducing players to represent their preferences truthfully. In his list of "Top Ten Open Research Questions" Camerer (2003) argues that many mechanisms can be cognitively too demanding to work in practice and that "experiments are an efficient way to 'test-bed' mechanisms and craft good theory" (p. 475). One of the aims of our study was to provide such a test for the Jackson–Sonnenschein mechanism. We found that in addition to its theoretical attractiveness, it is easily understood by subjects and hence they reap most

[33]The difference is also substantial. Taking the overall average across the groups for the two above measures yields 1939 and 1785, respectively, i.e. on average the two more honest subjects receive a 8.6% higher payoff than the two less honest subjects. Furthermore, we find that across all subjects the spearman rank-correlation between total payoffs and H_w is 0.31 and marginally misses significance ($p = 0.112$). Note, however, that this test ignores the dependence of observations within groups.

of the available efficiency gains. Alternative mechanisms such as the Clark–Groves mechanism are cognitively much more demanding which makes it substantially more difficult for experimental subjects to reach the levels of efficiency that they theoretically allow for.

In contrast, various forms of social interaction have produced truthful revelation to a much smaller degree, if at all. Only if the design suggests a linking of a limited number of problems in a straightforward way or if players can choose their partners, is there an effect on the honesty rates, which, however, translates into very small efficiency gains. That these efficiency gains are substantially smaller than those achieved by an exogenous budget does not come as a surprise, as the latter allows subjects to link all decisions, while any conditionally cooperative strategy can only link a subset of decisions. On the one hand, this shows the strength of the linking mechanism, on the other hand, this makes the latter a somewhat unfair benchmark for the social interaction treatments.

One might conclude that to arrive at an efficient outcome in such situations with private information, a central authority that enforces a budget is required. This might, however, be a premature conclusion. There are further aspects of social interaction that we have not investigated, but which might be more important outside the laboratory. For example, if players had an explicit punishment mechanism available, this might enable them to force each other to stick to a budget. Another important aspect of social interaction that we did not investigate here is communication, which could help to overcome the coordination problem. One might argue that given that we did not allow for communication the effect of competition alone is rather remarkable.

A further reason for the relatively low achieved level of cooperation is most likely that in the game we have studied, the outcome in the dominant–strategy equilibrium of the stage game may not be sufficiently miserable to get the players to try hard enough to overcome the problem. They have short–term incentives to overstate their preferences and even in the long run they obtain an acceptable, though inefficient outcome. We might see more creative approaches by the subjects if overstating of preferences resulted in zero or negative payoffs. In the study by Kaplan and Ruffle (2005), for example, the possible efficiency gains are substantially larger. This might be one reason why in their experiments subjects manage quite well to coordinate on efficient cut–off strategies.[34] While the main result of

[34] Another reason might be that the (private) signal is finer.

Kaplan and Ruffle does not agree with ours, there is an interesting similarity. We find that it does not help if ex–post information on true preferences is provided. Kaplan and Ruffle also find that this does not improve efficiency substantially.

To summarize, we observed that private information about preferences makes cooperation difficult, even in repeated interaction settings that enable subjects in many types of experiments to reap gains from cooperation. We also saw that the fact that information remains private ex–post does not appear to be the major problem, since changing this had virtually no effect. Instead, the crucial problems appear to be to coordinate on conditionally cooperative strategies if these can be based only on stochastic information and that it is relatively difficult for all parties to see the incentives to coordinate. We provide some evidence that the coordination problem is reduced if subjects decide upon several problems simultaneously, though this awaits more systematic investigation. Competition for partners is most effective in reducing these problems and enables endogenously formed pairs to cooperate. Hence competition has benefits beyond those traditionally identified in economics.

References

Camerer, Colin F. (2003). Behavioral Game Theory. Princeton, New Jersey: Princeton University Press.

Casella, Alessandra (2005). Storable Votes. Games and Economic Behavior, 51(2) 391–419.

Casella, Alessandra, Andrew Gelman, and Thomas R. Palfrey (2006). An Experimental Study on Storable Votes. Games and Economic Behavior, 57(1) 123–154.

Charness, Gary and Rabin, Matthew. (2002). Understanding Social Preferences with Simple Tests. Quarterly Journal of Economics, 117(3), 817–69.

Cooper, Russel W., Douglas V. DeJong, Robert Forsythe, and Thomas W. Ross (1996). Cooperation without Reputation: Experimental Evidence from Prisoner's Dilemma Games. Games and Economic Behavior 12(2), 187–218.

Coricelli, Giorgio, Dietmar Fehr, and Gerlinde Fellner (2004). Partner Selection in Public Goods Experiments. Journal of Conflict Resolution, 48(3), 356–378.

Engelmann, Dirk and Urs Fischbacher (2003). Indirect Reciprocity and Strategic Reputation Building in an Experimental Helping Game. CERGE-EI working paper No. 215.

Engelmann, Dirk and Strobel, Martin (2004). Inequality Aversion, Efficiency and Maximin Preferences in Simple Distribution Experiments. American Economic Review 94(4), 857–869.

Fischbacher, Urs (2007). Z-Tree, Zurich Toolbox for Readymade Economic Experiments. Experimental Economics 10(2), 171–178.

Hauk, Ester and Rosemarie Nagel (2001) Choices of Partners in Multiple Two-Person Prisoner's Dilemma Games: An Experimental Study. Journal of Conflict Resolution, 45(5), 770–793.

Hortala–Vallve, Rafael (2004). Qualitative Voting. Mimeo: LSE.

Hortala–Vallve, Rafael and Aniol Llorente–Saguer (2006). Qualitative Voting in Conflict Resolution Situations. Mimeo: Oxford University.

Huck, Steffen, Lünser, Gabriele K., and Tyran, Jean–Robert (2006). Competition Fosters Trust. ELSE working paper 235.

Jackson, Matthew and Hugo Sonnenschein (2007). Overcoming Incentive Constraints by Linking Decisions. Econometrica 75(1), 241–257.

Kaplan, Todd and Bradley J. Ruffle (2005). Which Way to Cooperate? Mimeo: University of Exeter.

Myerson, Roger B. and Mark A. Satterthwaite (1983). Efficient Mechanisms for Bilateral Trading. Journal of Economic Theory 28, 265–281.

Ule, Aljaž (2006). Cooperation with a Competition for Partners. Working Paper, University of Amsterdam.

A Payoffs from being honest, tough, and nice

An agent's expected payoff from being honest if the other is honest as well is

$$EU(h,h) = \frac{1}{2}\left(\frac{1}{2}\frac{1}{2}+\frac{1}{2}\right)2w + \frac{1}{2}\left(\frac{1}{2}\frac{1}{2}\right)w = \frac{7}{8}w \tag{4}$$

If both agents always report s or both always report w the decision is always taken by a flip of a coin and an agent's payoff would be

$$EU(s,s) = EU(w,w) = \frac{1}{2}\frac{1}{2}(2w) + \frac{1}{2}\frac{1}{2}w = \frac{3}{4}w \tag{5}$$

Hence, the relative efficiency gain of coordination on truthful behavior is $\frac{EU(h,h)-EU(s,s)}{EU(s,s)} = \frac{1}{6}$. However, deviation from honest behavior is profitable. The agent's expected payoff from always reporting s if the other is honest is

$$EU(s,h) = \frac{1}{2}\left(\frac{1}{2}\frac{1}{2}+\frac{1}{2}\right)2w + \frac{1}{2}\left(\frac{1}{2}\frac{1}{2}+\frac{1}{2}\right)w = \frac{9}{8}w \tag{6}$$

Thus, the relative incentive to lie if the other is honest is $\frac{EU(s,h)-EU(h,h)}{EU(h,h)} = \frac{2}{7}$, which is clearly higher than the efficiency gain from coordination on honest behavior. The expected payoff of being honest if the other always reports s is

$$EU(h,s) = \frac{1}{2}\frac{1}{2}(2w) + \frac{1}{2}(0w) = \frac{1}{2}w, \tag{7}$$

which yields a relative incentive to lie if the other is lying as well of $\frac{EU(s,s)-EU(h,s)}{EU(h,s)} = \frac{1}{2}$. Finally, if the other always states a weak preference, then being honest yields an expected payoff of

$$EU(h,w) = \frac{1}{2}(2w) + \frac{1}{2}\frac{1}{2}w = \frac{5}{4}w, \tag{8}$$

while the expected payoff of always stating a strong preference is

$$EU(s,w) = \frac{1}{2}(2w) + \frac{1}{2}w = \frac{3}{2}w, \tag{9}$$

and hence the relative incentive to lie if the other always states a weak preference is $\frac{EU(s,w)-EU(h,w)}{EU(h,w)} = \frac{1}{5}$. The expected payoff from always stating a weak preference if the other always states a strong preference is obviously zero while if the other states preferences truthfully, it is

$$EU(w,h) = \frac{1}{2}\frac{1}{2}\frac{1}{2}(2w) + \frac{1}{2}\frac{1}{2}\frac{1}{2}w = \frac{3}{8}w. \tag{10}$$

B Proofs of Propositions 1, 2, 3, 4, and 5

In this appendix we provide the proofs of our five propositions. In order to remind the reader of the particular equilibria we want to establish, we state the respective proposition prior to each proof.

Proposition 1 *In the infinitely repeated game with fixed matching between two players (as in FIX), the following constitutes a subgame–perfect Nash equilibrium. Players consider periods in (disjoint) blocks of two and adopt the following trigger strategy: state one strong and one weak preference (truthfully if possible, randomly otherwise) in a block of two periods if both players have done so in the past. If one player deviates, always state strong preferences for 16 periods, then go back to the pattern of one strong and one weak preference. The ex–ante (i.e. before preferences are drawn) expected payoff per period in this equilibrium is* $\frac{13}{16}$*. The ex–ante expected honesty rates are* $H_w = H_s = \frac{3}{4}$*.*

Proof First, deviating on the punishment path does not pay. If the other player sticks to the equilibrium strategy and states strong preferences for sixteen periods, there is nothing to be gained from stating any weak preferences in this part. Next, note that deviating to stating two weak preferences does obviously not pay and that deviating to stating two strong preferences pays most if one actually has two strong preferences. Thus we only need to consider this case. Then stating two strong preferences yields an expected payoff of $2 + \frac{1}{2} \cdot 2 = 3$ (since one project will be won for sure and the other will be assigned randomly). Stating one strong and one weak instead yields an expected payoff of 2. Thus the gain from deviating in the current block is 1. Now consider the loss in a punishment block compared to a "normal" block of two periods. If the player has either two strong or two weak preferences, the expected payoff is not affected, because in expectation he wins one of the projects, both in a punishment block and in a normal block. Punishment only affects the expected payoff if the player has one strong and one weak preference (for the same reason, why linking two decisions increases the efficiency only in this case, because it increases the probability of winning the project with the strong preference). In a normal block, the expected payoff in this case is (note that the player will state the strong preference for the truly strong preference, while the other player will state a strong preference with probability 1/2 in each case, so the project with the strong preference will be won with probability $\frac{3}{4}$ and the other with probability $\frac{1}{4}$.)

$$\frac{3}{4} \cdot 2 + \frac{1}{4} \cdot 1 = \frac{7}{4}$$

In a punishment block, the expected payoff when having one strong and one weak preference is

$$\frac{1}{2} \cdot 2 + \frac{1}{2} \cdot 1 = \frac{3}{2}$$

Since the probability to have one strong and one weak preference is 1/2, the difference in expected payoffs between a normal and a punishment block of two periods is thus $\frac{1}{2}\left(\frac{7}{4} - \frac{6}{4}\right) = \frac{1}{8}$. Thus if the punishment lasts for eight blocks, a player is indifferent between deviating or sticking to the equilibrium strategy.[35]

A player has with probability $\frac{1}{4}$ two strong preferences in a block of two periods, with probability $\frac{1}{2}$ one strong and one weak preference and with probability $\frac{1}{4}$ two weak preferences. In the second case, the expected payoff in equilibrium is $\frac{7}{4}$ (see above) and similarly in the first it is 2 and in the third it is 1. Hence the total ex-ante expected payoff from a block of two periods is

$$\frac{1}{4} \cdot 2 + \frac{1}{2} \cdot \frac{7}{4} + \frac{1}{4} \cdot 1 = \frac{13}{8}$$

Thus the ex-ante expected payoff per period is $\frac{13}{16}$.

To calculate the ex-ante expected honesty rates, let $p_w(x)$ denote the conditional probability of a weak preference to be in a constellation with x weak preferences (here within a block of two projects) and $h_w(x)$ the probability that a weak preference will be stated

[35]Obviously, extending the punishment path to more periods results in a strict preference to stick to the equilibrium strategy, but reduces the efficiency gains if punishment ever occurs. We ignore here that in the infinite game with discounting the expected payoffs are infinite anyway. Thus we treat the game from the perspective of a subject who treats the game as infinite in the sense that he does not consider a definite end from which to start backward induction, but is aware of the fact that the game has a final end and thus finite payoffs so that payoff comparisons are meaningful.

truthfully in such a constellation. Then $p_w(1) = p_w(2) = \frac{1}{2}$, $h_w(1) = 1$ and $h_w(2) = \frac{1}{2}$. This yields $H_w = \sum p_w(x)h_w(x) = \frac{1}{2} + \frac{1}{2}\frac{1}{2} = \frac{3}{4}$. The situation for strong preferences is symmetric, so $H_s = \frac{3}{4}$.

□

PROPOSITION 2 *In the infinitely repeated game with random link formation in a fixed group of four players (as in RLK), the following constitutes a subgame–perfect Nash equilibrium. Players consider one period at a time and adopt the following trigger strategy: for one project, state a strong preference, for two or three projects, state one weak preference and for four projects, state two weak preferences if all players have followed this pattern in the past. State preferences as truthfully as possible given these restrictions. If one player deviates, always state strong preferences for all projects for 19 periods, then go back to the above pattern. The ex–ante expected payoff per period in this equilibrium is $\frac{695}{432}$. The ex–ante expected honesty rates are $H_w = \frac{127}{216} \approx 58.8\%$ and $H_s = \frac{61}{72} \approx 84.7\%$.*

PROOF First, deviating on the punishment path does not pay. If the other players stick to the equilibrium strategy and state strong preferences for 19 periods, there is nothing to be gained from stating any weak preferences in this part.

The fact that the equilibrium strategy requires different probabilities of stating a strong preference conditional on the number of projects might actually suggest that a player should condition his distribution of weak and strong stated preferences on the number of projects the respective partners are involved in. Note, however, that independent of the stated preference of the partner, stating a strong instead of a weak preference always increases the probability of winning the project by 1/2, so at least for risk–neutral players, the expected stated preference and hence the number of projects of the partner is irrelevant for how to optimally allocate the permitted strong stated preferences.

Next, note that deviating to stating more weak preferences than required by the equilibrium strategy does obviously not pay and that deviating to stating more strong preferences than prescribed by the equilibrium strategy pays most if one has four projects and the true preference for all these projects is strong. Thus we only need to consider this case. For each project with a strong true preference, stating a strong instead of a weak preference increases the probability of winning this project by 1/2. Thus stating four instead of two strong preferences if a player actually has four strong preferences, increases the expected payoff by $2 \cdot \frac{1}{2} \cdot 2 = 2$.

Now consider the loss in a punishment period compared to a "normal" period. To calculate the expected payoff in a normal period note first that the probabilities to have 4, 3, 2, and 1 project(s) are $\frac{1}{27}$, $\frac{2}{9}$, $\frac{4}{9}$, and $\frac{8}{27}$, respectively. Straightforward, but tedious calculation shows that if a player has four projects, the expected payoff given the equilibrium strategies of the other players in a normal period is $E_n(4) = \frac{41}{16}$. Similarly, for three, two and one projects $E_n(3) = \frac{119}{48}$, $E_n(2) = \frac{35}{24}$, $E_n(1) = \frac{17}{16}$. Thus the ex–ante expected payoff in a normal period is

$$\frac{1}{27}E_n(4) + \frac{2}{9}E_n(3) + \frac{4}{9}E_n(2) + \frac{8}{27}E_n(1) = \frac{1}{27}\frac{41}{16} + \frac{2}{9}\frac{119}{48} + \frac{4}{9}\frac{35}{24} + \frac{8}{27}\frac{17}{16} = \frac{695}{432}$$

In a punishment period all players state strong preferences, so any project will be won with probability $\frac{1}{2}$ and so since its expected value is $\frac{3}{2}$, each project yields an expected payoff of $\frac{3}{4}$. Furthermore, the expected number of projects is 2, so the ex–ante expected payoff in a punishment period is $\frac{3}{2}$. Thus the expected loss from a punishment period compared to a normal period is $\frac{695}{432} - \frac{3}{2} = \frac{47}{432}$. Since 19 is the smallest integer t such that $\frac{47}{432}t > 2$, not deviating from the equilibrium strategy pays if the punishment lasts at least 19 periods.

To calculate the ex–ante expected honesty rates, let $p_w(x|y)$ denote the conditional probability of a weak preference to be in a constellation with y projects and x weak preferences among these and $h_w(x|y)$ the probability that a weak preference will be stated truthfully in such a constellation. Then we obtain $p_w(1|1) = \frac{4}{27}, h_w(1|1) = 0, p_w(1|2) = \frac{2}{9}, h_w(1|2) = 1, p_w(2|2) = \frac{2}{9}, h_w(2|2) = \frac{1}{2}, p_w(1|3) = \frac{1}{12}, h_w(1|3) = 1, p_w(2|3) = \frac{1}{6}, h_w(2|3) = \frac{1}{2}, p_w(3|3) = \frac{1}{12}, h_w(3|3) = \frac{1}{3}, p_w(1|4) = \frac{1}{108}, h_w(1|4) = 1, p_w(2|4) = \frac{1}{36}, h_w(2|4) = 1, p_w(3|4) = \frac{1}{36}, h_w(3|4) = \frac{2}{3}, p_w(4|4) = \frac{1}{108}, h_w(4|4) = \frac{1}{2}$. Then $H_w = \sum p_w(x|y)h_w(x|y) = \frac{127}{216}$. Similarly, with $p_s(x|y)$ and $h_s(x|y)$ defined in a corresponding way, we obtain $p_s(1|1) = \frac{4}{27}, h_s(1|1) = 1, p_s(1|2) = \frac{2}{9}, h_s(1|2) = 1, p_s(2|2) = \frac{2}{9}, h_s(2|2) = \frac{1}{2}, p_s(1|3) = \frac{1}{12}, h_s(1|3) = 1, p_s(2|3) = \frac{1}{6}, h_s(2|3) = 1, p_s(3|3) = \frac{1}{12}, h_s(3|3) = \frac{2}{3}, p_s(1|4) = \frac{1}{108}, h_s(1|4) = 1, p_s(2|4) = \frac{1}{36}, h_s(2|4) = 1, p_s(3|4) = \frac{1}{36}, h_s(3|4) = \frac{2}{3}, p_s(4|4) = \frac{1}{108}, h_s(4|4) = \frac{1}{2}$ and $H_s = \sum p_s(x|y)h_s(x|y) = \frac{61}{72}$.

$\square$

Proposition 3 *In the finitely repeated two-stage game where players first choose a partner for the project originating with them and then state preferences for all their projects (as in CMP), the following constitutes a subgame–perfect Nash equilibrium. Let the number of periods be n. Players consider one period at a time and adopt the following trigger strategy: in the first stage, two pairs form endogenously, that is two players choose each other and hence share two projects. In the second stage, both players in each pair state one strong and one weak preference. Both players continue in this fashion until period $n-2$ if they have done so in the past. In period $n-1$, they will state two strong preferences if they have indeed two strong preferences but again one weak and one strong otherwise. In period n, if no player stated two strong preferences before, they still choose each other but state two strong preferences. If player i deviates by stating preferences other than one strong and one weak in any period, then for the rest of the game, player j (the "partner" of i) chooses another player, while player i continues to choose j and they both state strong preferences. If a player is chosen by a player from a different pair, he states a strong preference for the resulting project. The ex–ante expected payoff per period in this equilibrium is $\frac{13}{8}$. The ex–ante expected honesty rates are $H_w = H_s = \frac{3}{4}$.*

Proof

We start in the last period, n. Since there is no future, there is no reason to state a weak preference, so no player will deviate in the second stage of period n. Since all players will state strong preferences, it is irrelevant which partner is chosen, so there is no strict incentive to deviate from the equilibrium partner choice in the first stage of period n. Now consider period $n-1$. Obviously, the only plausible deviation is to state two strong preferences. This increases the probability to win the second project by $\frac{1}{2}$, so the expected gain is 1 if the player actually has two strong preferences and $\frac{1}{2}$ if he has at most one strong preference (since he would in equilibrium state a strong preference for his true strong preference if he has one). If he states two strong preferences, he loses a project in period n.[36] The expected payoff from this project is (since both players will state a strong preference, it is won with probability $\frac{1}{2}$) $\frac{1}{2}(\frac{1}{2}\cdot 2+\frac{1}{2})=\frac{3}{4}$. Hence it pays to state two strong preferences in period $n-1$ if and only if a player has two strong preferences. Now consider period $n-2$. Again, the gain from stating two strong preferences if the true preferences are indeed strong is 1. As a consequence, the player would have only one project (unless he is chosen by a player from the other pair, but that happens independent of the behavior of himself or his partner, so can be ignored) in periods $n-1$ and n, with strong preferences stated by both players and hence expected payoffs of $\frac{3}{4}$ in both periods. Thus the payoff gained in $n-2$ from deviating plus the resulting payoff from $n-1$ and n is $1+\frac{3}{4}+\frac{3}{4}=\frac{5}{2}$. This has to be compared with the expected payoff from periods $n-1$ and n if he does not deviate in $n-2$. We have to take into account that both i and j would state two strong preferences (which they have with probability $\frac{1}{4}$) truthfully in $n-1$. If i has two strong preferences in $n-1$, his expected payoff is (the first $\frac{3}{4}$ being the expected payoff from period n as he will only have one project then)

$$\frac{3}{4}+\frac{3}{4}(2+\frac{1}{2}\cdot 2)+\frac{1}{4}(\frac{1}{2}\cdot 2+\frac{1}{2}\cdot 2)=\frac{3}{4}+\frac{3}{4}\cdot 3+\frac{1}{4}\cdot 2=\frac{7}{2}.$$

If i has two weak preferences in $n-1$, his expected payoff is (the first $\frac{3}{2}$ being the expected payoff from period n as he will have two projects then)

$$\frac{3}{2}+\frac{3}{4}(\frac{1}{4}\cdot 1+\frac{3}{4}\cdot 1)+\frac{1}{4}(\frac{1}{2}\cdot 1)=\frac{3}{2}+\frac{3}{4}+\frac{1}{8}=\frac{19}{8}.$$

Finally, if i has one strong and one weak preference in $n-1$, his expected payoff from periods $n-1$ and n is

$$\frac{3}{2}+\frac{3}{4}(\frac{1}{4}\cdot 1+\frac{3}{4}\cdot 2)+\frac{1}{4}(\frac{1}{2}\cdot 2)=\frac{3}{2}+\frac{21}{16}+\frac{1}{4}=\frac{49}{16}.$$

Thus his overall expected payoff from $n-1$ and n if he does not deviate in $n-2$ is

$$\frac{1}{4}\cdot\frac{7}{2}+\frac{1}{4}\cdot\frac{19}{8}+\frac{1}{2}\cdot\frac{49}{16}=3>\frac{5}{2}$$

Thus it does not pay to deviate from stating one strong and one weak preference in period $n-2$ even if the true preferences are both strong. If the horizon is longer, the total losses

[36] Note that if j deviates in $n-1$, he will still choose i in n (unless i deviates as well), so when considering whether to deviate, i does not have to take into account the probability that j deviates.

from losing the partner in the remaining periods is obviously even higher, so it does not pay to deviate in any of the previous periods.

Finally note that since each player has two projects in each period of this equilibrium and that these two projects are linked and budgeted as in a block of two projects in the above equilibrium of FIX, the ex–ante expected payoff in each period is equal to the ex–ante expected payoff in a block of two periods in that equilibrium and hence equal to $\frac{13}{8}$ and the ex-ante expected honesty rates are $H_w = H_s = \frac{3}{4}$.

□

PROPOSITION 4 *In the infinitely repeated game with fixed matching between two players and ex–post information about the intensity of preferences (as in FXI), the following constitutes a subgame–perfect Nash equilibrium. Players state their preferences truthfully if both players have done so in the past. If it becomes known at the end of a period that a player misrepresented his preference, both players state always strong preferences for four periods and then go back to honest representation. The ex–ante expected payoff per period in this equilibrium is $\frac{7}{8}$, and the ex–ante expected honesty rates are $H_w = H_s = 1$.*

PROOF

Obviously, there is no incentive to deviate in a punishment period and the only deviation to consider in a normal period is to state a strong preference when the true preference is weak. Since this raises the probability of winning by $\frac{1}{2}$, the expected gain from the deviation is $\frac{1}{2}$. The ex–ante expected payoff in a normal period is $\frac{7}{8}$, while that in a punishment period is $\frac{3}{4}$. Thus deviating does not pay if the punishment lasts for at least four periods. The expected honesty rates follow immediately from the fact that preferences are stated truthfully in equilibrium.

□

PROPOSITION 5 *In the infinitely repeated game with fixed matching between two players and four projects in each period (as in F4P), the following constitutes a subgame–perfect Nash equilibrium. Players play the following trigger strategy: state two strong and two weak strategies (as truthfully as possible, randomly otherwise) in each period if both players* *have done so in the past. If one player deviates, always state strong preferences for seven periods, then go back to the pattern of two strong and two weak preferences. The ex–ante expected payoff per period in this equilibrium is $\frac{53}{16}$. The ex–ante expected honesty rates are $H_w = H_s = \frac{13}{16}$.*

PROOF

Obviously, deviating in a punishment period does not pay and deviating in a normal period pays most when the true preferences are all strong and a player would deviate to state only strong preferences. Thus we only need to consider this case. By deviating from stating a weak to stating a strong preference, the probability of winning this project is increased by $\frac{1}{2}$ and hence the expected gain from this deviation is 2.

The expected payoff from each project in a punishment period is the stage–game Nash–equilibrium payoff of $\frac{3}{4}$, so the total ex–ante expected payoff in a punishment period is 3. To calculate the ex–ante expected payoff in a normal period, note that with probability $\frac{1}{16}$ a player has four strong preferences, and with probabilities $\frac{1}{4}$, $\frac{3}{8}$, $\frac{1}{4}$ and $\frac{1}{16}$ he has three, two, one, or no strong preference. The respective expected payoffs in a normal period are 4, $\frac{15}{4}$, $\frac{7}{2}$, $\frac{11}{4}$, and 2. Thus the ex-ante expected payoff in a normal period is

$$\frac{1}{16} \cdot 4 + \frac{1}{4}\frac{15}{4} + \frac{3}{8}\frac{7}{2} + \frac{1}{4}\frac{11}{4} + \frac{1}{16} \cdot 2 = \frac{53}{16}.$$

Thus a punishment period leads to a loss in expected payoff of $\frac{5}{16}$ and hence if the punishment lasts for at least seven periods, deviating from the equilibrium path does not pay.

With $p_w(x)$ and $h_w(x)$ as above we obtain $p_w(1) = p_w(4) = \frac{1}{8}$, $p_w(2) = p_w(3) = \frac{3}{8}$, $h_w(1) = h_w(2) = 1$, $h_w(3) = \frac{2}{3}$, $h_w(4) = \frac{1}{2}$, so $H_w = \sum p_w(x)h_w(x) = \frac{1}{8} + \frac{3}{8} + \frac{3}{8}\frac{2}{3} + \frac{1}{8}\frac{1}{2} = \frac{13}{16}$. For strong preferences the situation is symmetric, so $H_s = \frac{13}{16}$.

□

Instructions for Competition Treatment (not for Publication)

General Instructions

You are taking part in an experiment on decision-making. If you read the following instructions carefully, you can - depending on the decisions you and other participants of this experiment will make - influence your own earnings as well as the earnings of the other participants. It is, therefore, important that you pay attention to the instructions given below. These instructions are the same for all participants.

The instructions distributed are intended for your personal information only. Please do not talk to any of the other participants for the duration of the experiment. Please address questions you might have to us directly.

The experiment is divided into **periods**. In each period you will face the same situation that is described below. During this experiment we will calculate your earnings directly in Pence. Your income from each period, as well as your cumulative earnings, will, therefore, be stated in Pence. At the end of the experiment we will pay you your earnings in cash, at the known rate of 1£=100 Pence.

This experiment will last for 40 periods.

Whom you interact with

Throughout the experiment you form a **group of four** with the same three other participants of the experiment. You will be able to identify previous choices of the others since each group member (including you) is assigned a "name" (A-D) and will keep this "name" throughout the experiment. After the experiment, however, you will not be able to infer which participants you were actually interacting with as the "names" and group composition will not be revealed.

Each of the 40 periods is divided into two **stages**.

In the **first** stage, each person in your group chooses a link to one other group member. Each two group members that are linked in this way will be involved in a joint project. This means that in each period you will be involved in one to four projects, depending on how many links you have gotten in that period. You will be involved in at least one project (since you choose one link to another person) and in at most four projects (if all the other three members of your group choose you). In each new period, the links are chosen anew.

In the **second** stage, each two participants involved in a joint project will have to make a joint decision. Thus, you are involved in one to four independent decisions. Your payment and the payment of the other participant involved in a particular project will depend on the decision that is taken according to the rules described below on the basis of your actions.

If you are involved in more than one project, it is important to note that - within each period - the decisions are completely independent across projects. This means that in each period your actions taken with respect to one project will have no effect whatsoever on the payments you get from another project. In particular, you can be involved in two decisions with one other participant in one period, namely if you choose her or him and he or she chooses you. Even in this case, these two projects are independent, that is, the actions taken concerning one project do not influence the payments from the other.

What are you paid?

Each project can be executed in two different versions. The two participants involved in a project always disagree on the version of the project that should be executed and this is always known to both of them.

For each project you are involved in, you will get a positive payment if your preferred version is chosen. If the other's preferred version is chosen, you will get nothing.

Each project can be more or less important to you. We will call the importance the project has for you its **priority**. Hence for each single project, the priority that the project has for you can either be **high** (in which case you would get 60 Pence if your preferred version is chosen), or **low** (in which case you would get 30 Pence if your preferred version is chosen). Both, a low and a high priority, are equally likely. For each project you are involved in, these priorities are independently drawn. This means that whether your priority for one project is high or low does not influence the probability that it is high for another project you are involved in, either in the current or any future period.

For each project, you learn the priority you assign to the project before you are required to make a decision. The other participant will not be informed about your priority. He or she only knows that it can be high or low, and that both are equally likely.

Similarly, you do not know the priority the other participant assigns to the project. All you know is also that it can be high or low, and that both are equally likely. The draws for both of you are independent. This means no matter whether your priority is high or low, it is always equally likely that the other player's priority will be high or low.

Thus, for each project, the two of you involved in the decision do not know for whom it is more valuable if his or her preferred version of the project becomes accepted, or whether both of you value it equally. Note that for each project only one participant involved can get a positive payment. Who that is depends on the version of the project that is chosen.

How is the decision taken?

In each period, after the decisions in the first stage have been made, you are informed about all the links in your group. Hence you will know how many projects you are involved in and also the "name" of the respective other participant in each project. For each project you are involved in, you will then observe the priority you assign to that project. The priority can be different for different projects you are involved in.

For each project, after you have observed the priority you assign to the project, you and the other participant involved in that project decide on the version that should be chosen.

Both of you will be asked to state a priority for the project. This stated priority does not have to be the same as your true priority.

For each project, the version, for which the higher priority is stated will be chosen. That is, if either you or the other participant states a high priority and the other a low priority, the preferred version of the person who stated the high priority will be chosen. If, however, both of you state the same priority, the decision will be taken randomly whereby both versions will equally likely be chosen. This is illustrated in the following table.

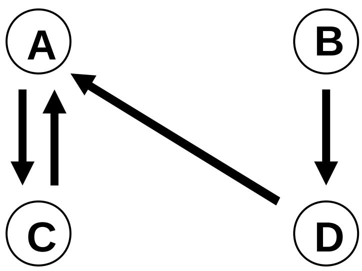
		The other participant states a	
		High priority	Low Priority
You state a	High priority	Random decision, each version with probability 50%	Your preferred version is chosen
	Low priority	The other participant's preferred version is chosen	Random decision, each version with probability 50%

If your preferred version of the project is chosen, you will get 60p or 30p, depending on your true priority and the other participant will receive nothing. If the other participant's preferred version is chosen, he or she will receive 60p or 30p, depending on his or her true priority and you will receive nothing. Your possible payments are illustrated in the following table

		The chosen version is	
		Your preferred	The other's preferred
Your priority is	High	You get 60p	You get 0p
	Low	You get 30p	You get 0p

An example

Consider the following example: assume you are participant A. You choose a link to C. B chooses D, C chooses you and D also chooses you. Hence you are involved in **three** projects, one with D and **two** with C (because the link going from you to C and the one going from C to you correspond to two independent projects). B is involved in only one project (with D), C is involved in two projects (both with you) and D is also involved in two projects, one with B and one with you.



Now assume that for the first project with C you have a high priority and for the second project with C you have a low priority. For the project with D you also have a low priority. Assume that for the first project with C you state a high priority and C states a low priority. Hence your preferred version of this project is implemented and you receive 60p. Furthermore, assume that for the second project with C you state a low priority and C states a high priority. Hence the version that C prefers will be chosen, so you receive nothing for the second project. For the project with D assume that you both state a high priority and that the randomly chosen version is the one you prefer. Hence (since your real priority was low), you receive 30p. Thus in this period you receive 90p in total.

Information you receive

At the end of each period, you will be presented three screens that contain information about the current and previous periods.

The first screen will contain all information that is relevant for your earnings in the current period. For each of the projects you are involved in, you will be able to observe the stated priority of the other participant, your stated priority, and the decision that has been taken. You will also be informed about your payment from each project, your total payment from the current period and your cumulative earnings up to that period.

The second screen shows for all four projects in your group the priorities stated by both participants.

On a third screen, you will get a summary of what happend up to the current period. This screen contains the stated priorities by all members in your group of four (A, B, C, and D) in all preceding periods. For each participant, it will simply list the stated priorities, but not indicate who was involved in a project with whom. Every other participant in your group will also be presented this table.

Summary

- The experiment will last for 40 periods.
- Throughout those 40 periods you will form a group of four with the same three other participants.
- In the first stage of each period, each group member chooses a link to one other group member.
- In the second stage, each two group members that are linked engage in a joint project.
- In each period, you may be involved in one to four projects.
- In each period, actions you take with respect to one project have no payoff consequences on any other project in this period.
- For each project, you and the other person disagree on the version of the project that should be chosen.
- For each project you are involved in, you observe your priority of the project, but not the other's.
- For each project, you and the other participant are both asked to state a priority of the project. This does not have to be equal to your true priority.
- For each project, the version for which the higher priority is stated is chosen.
- If both players state equal priorities, the version will be chosen randomly.
- For each project you are involved in, you will be paid according to your true priority if your preferred version is chosen, otherwise you will receive nothing.
- At the end of each period you receive all information that is payoff relevant for you and information on the actions taken by all members of your group.

Please answer the following questions before we start the experiment

1. Are the other participants whom you form a group with in period 2 the same as in period 1?

2. Assume you are person B. Let the links in your group be as follows: A chooses D, B chooses A, C chooses A, D chooses B. How many projects are you involved in? How many projects are the other participants involved in?

3. Assume that for one project both you and the other participant state a low priority. Whose preferred version will be chosen?

4. Assume that for one project your priority is low and the other participant's priority is high. If you state a low priority and the other participant a high priority, what are the payments for you and the other participant from this project?

5. Assume that for one project your priority is low and the other participant's priority is high. If you state a high priority and the other participant a low priority, what are the payments for you and the other participant from this project?

6. Assume that for one project your priority is high and the other participant's priority is low. If you both state a high priority and the random procedure picks the other participant's preferred version, what are the payments for you and the other participant from this project?

7. Assume that in the current period you are involved in 3 projects. For the first project, your priority is high and your preferred version is chosen. For the second project, your priority is high and the other's preferred version is chosen. For the third project, your priority is low and your preferred version is chosen. What is your total payment in this period?

Diskussionspapiere 2008
Discussion Papers 2008