No. 22

UNIVERSITY OF COLOGNE

WORKING PAPER SERIES IN ECONOMICS

OVERCOMING INCENTIVE CONSTRAINTS? THE (IN-)EFFECTIVENESS OF SOCIAL INTERACTION

DIRK ENGELMANN
VERONIKA GRIMM

Department of Economics
University of Cologne
Albertus-Magnus-Platz
D-50923 Köln
Germany

http://www.wiso.uni-koeln.de

Overcoming Incentive Constraints? The (In-)effectiveness of Social Interaction *

Dirk Engelmann[†] Veronika Grimm[‡]

Royal Holloway, University of London University of Cologne

February 1, 2006

Preliminary Version

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. Applying the linking mechanism suggested by Jackson and Sonnenschein (2005) captures nearly all achievable efficiency gains. Repeated interaction leads to significant gains in truthful representation and efficiency only if 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 seminar participants at the University of Alicante, Tinbergen Institute, the University of Hannover, the University of Magdeburg, the University of Karlsruhe, Dundee University and the University of St. Andrews as well as participants at the ESA meetings in Montreal and Alessandria 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 Cologne, Staatswissenschaftliches Seminar, Albertus-Magnus-Platz, D-50923 Köln, Germany, vgrimm@uni-koeln.de

1 Introduction

In a recent paper Jackson and Sonnenschein (2005) 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 parties. In the course of the negotiations, agreements on a variety of topics have to be reached. From a social point of view, for each single decision it would be desirable if the party succeeded that cares more about the issue.[1] 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 at 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 being decided they could achieve a more favorable budget, which complicates the ex-ante

---

[1]We assume for simplicity of exposition that the parties' preferences coincide with those of their voters. Furthermore, we assume that the utilities of voters are comparable and that the number of voters for each party is about equal. Otherwise, the smaller party's preference should only be implemented if it cares substantially more about the issue than the larger party.

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

agreement on institutions that enable budgeting.

However, in practice economic agents 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 on the two sides of the market, as we will show in our paper.

In this paper, we compare the effectiveness of the exogenous budgets 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 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 (2003) 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 (2004a) 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. In Hortala-Vallve (2004b) he presents experimental support that subjects vote in accordance with the equilibrium predictions and that qualitative voting achieves the efficient outcome significantly more often than majority voting. 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 Tyran, Huck, and Ruchala (2005). In line with our results, they find that competition increases trustworthiness beyond the level achieved through reputation building alone.[3]

2 An Experiment on Linking Decisions

In section 2.1 we present a slightly modified version of one of Jackson and Sonnenschein's (2005) examples (in order to illustrate their point), and introduce two experimental treatments 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

[3] Coricello, 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.

[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 $s = 2w$.

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 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

[5]In case both players might prefer the same version with some 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.

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 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 (2005) show that linking more decisions helps further, and in the limit their mechanism leads to full Pareto efficiency.

As a first step of our experiment, 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 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.

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?

In the following we will argue that 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. 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. Obviously, it is required that decisions are sequential for a player to be able to reciprocate violations of the budget,[7] whereas an exogenous budget can be applied to several decisions that are made simultaneously as well as to those made sequentially.

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

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

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 him to reject the hypothesis that this sequence is random at some pre-determined level. This already points to a problem of endogenizing budgets. In order to know his own budget, a player needs to know the other player's strategy.[8]

In order to illustrate the possible benefits from linking decisions (either through exogenous budgets or through conditionally cooperative behavior), we now present the possible efficiency gains from honest behavior as compared to (stage-game Nash-)equilibrium play. This is what the agents could distribute among themselves if they coordinated on truthful behavior. 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. The expected payoffs are displayed in Table 3, for the computations see the appendix.

$EU(x,y)$		other player: $y =$		
		h	s	w
me: $x =$	h	$\frac{7}{8}w$	$\frac{1}{2}w$	$\frac{5}{4}w$
	s	$\frac{9}{8}w$	$\frac{3}{4}w$	$\frac{3}{2}w$
	w	$\frac{3}{8}w$	0	$\frac{3}{4}w$

Table 3: Payoffs from being honest, tough, and nice.

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

[8]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 outperform exogenous ones.

the other player always reports a strong preference, doing the same even increases the player's payoff by 50 % compared to honest behavior.

As we argued above, social interaction might allow players to endogenize a budget for their representation of preferences and thus to overcome the incentive constraints. The second aim of our study is 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. Specifically, 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 round bidders 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.

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 groups. 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.[9] 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 feedback 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 the truth and trying to build a reputation of almost always giving in, which in turn could lead to inefficiencies. The expected payoff from doing so, however, is rather low: If the partner is being honest, we have that $EU(w,h) = \frac{3}{8}w$ (compare Table 3), whereas the expected payoff is zero, if the other always reports a strong preference. Thus, always reporting a weak preference would not even pay if the player doubled the number of his honest partners.

[9] 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.

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 stage game if matching is random and players cannot identify players they interact(ed) with. Obviously, repeated interaction, the observability of past actions, and the possibility to select interaction partners affect the nature of equilibria of the game. All games have Nash equilibria that involve cooperative behavior. Note however, that the threats needed to establish cooperation are not credible in the treatments FIX and RLK.

An important observation is that in treatment CMP the possibility to choose partners establishes pair formation and budgeting (for example, but not exclusively with a one–weak–one–strong budget per period) as a subgame–perfect Nash–equilibrium even for the finitely repeated game. Such an equilibrium requires a strategy to abandon the partner for the rest of the game (or at least sufficiently long) if he violates his budget. The reason why cooperation can be established is that the expected gain from being involved in another project in the last period is higher than the expected gain from overstating a weak preference in the second to last period.[10] Hence each player in such an endogenously formed pair would, if he has at least one weak preference in the second to last round, state this truthfully. In case he has only two strong preferences, the expected gain on stating these truthfully[11] is higher than the expected gain from another project in the last period, so he will state his strong preferences truthfully and lose his partner in the last period. If the horizon is longer than one period, the total losses from losing the partner in all of these periods are higher than the immediate gain of stating both preferences as strong (even if they are) and hence a player will stick to his one–strong–one–weak budget.[12]

[10]Since in the last period, each player will state only strong preferences, a project will be won with probability $\frac{1}{2}$, so the expected gain from another project is $\frac{1}{2}(\frac{1}{2}2w + \frac{1}{2}w) = \frac{3}{4}w$ whereas stating a strong instead of a weak preference increases the probability of receiving a payoff by $\frac{1}{2}$ and hence increases the expected payoff by $\frac{1}{2}w$.

[11]As above we see that it is $\frac{1}{2}2w = w$.

[12]If there are two (or more) periods left, a player not only loses the project that originates from his partner, but his partner would also have no incentive anymore to state his preferences truthfully on the other project in the second to last period. So the expected gains from stating a second strong preference are, as above, w, and the expected payoff from the one project in each of the remaining two periods is

Even though calculating the subgame perfect Nash equilibrium is certainly demanding for the subjects, its logic is fairly obvious. Hence an important reason why subjects may state a weak preference more frequently in treatment CMP may well be that the opportunity to attract partners clearly provides additional incentives to *appear* honest. It can not only possibly trigger reciprocal honesty but also attract partners and hence increase the number of projects, which directly increases the expected payoffs.

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 section 2.1. In the experiment, a weak (strong) preference corresponded to a payoff of 30 (60) Pence in the case the desired version of the project was chosen. In each period, 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, 1999) 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 4 for details.

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

---

$2\frac{3}{4}w$. In contrast, sticking to the budget now leads to an expected payoff from the last two periods of at least (this assumes that if the other player violates the budget in the second to last period he will also switch to another partner in the last period, which is not required, otherwise the expected payoff is higher) $\frac{1}{4}[\frac{3}{4}(w+\frac{3}{2}w)+\frac{1}{4}(\frac{1}{2}w+\frac{3}{4}w)]+\frac{1}{2}[\frac{3}{4}(\frac{3}{4}2w+\frac{1}{4}w+\frac{3}{2}w)+\frac{1}{4}(\frac{1}{2}2w+\frac{3}{4}w)]+\frac{1}{4}[\frac{3}{4}(2w+\frac{1}{2}2w+\frac{3}{4}w)+\frac{1}{4}(2w+\frac{3}{4}w)] = \frac{183}{64}w > \frac{5}{2}w$. So it does not pay to state two strong preferences in the third to last period.

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

Table 4: Number of subjects and independent observations for the different experimental treatments

(including reading the instructions, answering a post-experimental questionnaire and receiving payments).[13] Average earnings ranged from 9.01 (treatment FXI explained below) to 18.64 (competition treatment) with an overall average of 13.13.[14]

3 Evaluation of the Data - Preliminaries

3.1 Relevant Measures of Experimental Behavior

There are two important aspects of the data 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

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

[13]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.

[14]Note that the large differences in average payoffs are not due to much more successful cooperation, but primarily occur because there are twice as many projects per subject in RLK and CMP than in the other treatments.

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

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

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

where Equal_p is a dummy that is 1 if both bidders 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 possible 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,

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

Note that if bidders 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 $E[E] = \frac{1}{2}$.

3.2 Hypotheses

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

   In treatment I, random matching (RAN), subjects are expected to overwhelmingly state a strong preference and that this rate approaches 100%. Expected Efficiency should hence be close to the minimum.

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

   In treatment II, exogenous budgets (EXO), subjects are expected to represent their preferences overwhelmingly honestly, in particular after some initial learning, as long as their behavior is not constrained by the budget. Efficiency should be close to the achievable maximum. Towards the end, honesty rates should decline, because subjects are either constrained if they have spent their budget or have some free budget.

3. $H_w(RAN) < H_w(FIX), H_w(RLK);\ H_w(FIX), H_w(RLK) < H_w(EXO)$,

   $E[E](RAN) < E[E](FIX), E[E](RLK);\ E[E](FIX), E[E](RLK) < E[E](EXO)$

   In the stable partnerships (FIX) and random links (RLK) treatments honesty rates should be higher than in RAN as subjects might realize gains from linking decisions and apply some stochastic conditionally cooperative strategies. Due to coordination problems and imperfect opportunities for monitoring behavior, honesty should be lower than in EXO. This should translate into efficiency higher than in RAN and lower than in EXO.

4. $H_w(CMP) > H_w(RLK),\ H_s(CMP) < H_s(RLK)$,

$E[E](CMP) > E[E](RLK)$.

In the competition treatment (CMP) there are additional incentives to appear honest as this might attract partners. Hence subjects are expected to state more often a weak preference than in RLK which differs only by the formation of pairs. This has two consequences. One the one hand, players should be more honest in CMP than in RLK if they have a weak preference. On the other hand, if they have a longer streak of strong preferences, they might feel forced to represent their preferences as weak, i.e. they might "lie down" in order to prevent the impression that they are "lying up". While the first effect has positive impact on efficiency, the second has negative impact. We would expect the first effect to dominate, resulting in higher expected efficiency.

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.

treatment	H_w	H_s
Random Matching RAN	7.0%	97.2%
Exogenous Budgets EXO	85.7% (88.7%)	84.9% (89.3%)

Table 5: Share of truthfully represented weak and strong preferences for RAN and EXO. The numbers in parentheses for EXO corrects 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.[15]

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 down". Even if we correct for this, however, the share of truthfully represented strong preferences rises to only 89.3%. Following statements in the post–experimental questionnaire, 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 only 4 out of 24 players always report their preferences truthfully, even after correcting for forced and free lies). 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 ($p = 0.1$).

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 overwhelm-*

[15]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.

ingly play their dominant strategy to state a strong preference.

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.

treatment	H_w	H_s
Random Matching RAN	7.0%	97.2%
Exogenous Budgets EXO	85.7% (88.7%)	84.9% (89.3%)
Stable Partnerships FIX	11.6%	97.9%
Random Links RLK	12.8%	98.2%
Competition CMP	30.9%	93.8%

Table 6: Share of truthfully represented weak and strong preferences for FIX, RLK, and CMP.

The honesty rates in FIX and RLK do not differ substantially or significantly from those in RAN. Essentially, it appears that simply 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.

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, $p < 5\%$).

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).*

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.

treatment	Expected Efficiency	Exp. Eff. Periods 1-10
Random Matching RAN	52.4%	58.4%
Exogenous Budgets EXO	87.2% (89.1%)	91.2%
Stable Partnerships FIX	55.0%	55.8%
Random Links RLK	55.9%	59.4%
Competition CMP	62.1%	70.3%

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 rightmost column shows the expected efficiency in Periods 1 to 10.

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 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.[16]

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 efficiency 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 that this translates into higher expected efficiency (hypothesis 4). Contrary to hypothesis 3, fixed matching in pairs

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

or groups of four and the possibility of reputation building has a significant effect neither on honesty rates nor on expected efficiency. Based on those (partly surprising) results, in the next section, we try to isolate what drives successful cooperative behavior in our environment of two sided private information.

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 Tyran, Huck, and Ruchala, 2005).

We will now address possible explanations for the observed failure of cooperation in FIX and RLK and the relatively low efficiency gain 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 punish. 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 (our 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 this, they can choose each other and form an endogenous 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.[17]

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 that the other player had. 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 round, F4P. The treatment is identical to "Stable Partnerships" (FIX), except for the fact that each pair decided on four independent projects each round. 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.[18] The payoffs for strong (weak) preferences

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

[18] Of course, they could in principle still link more decisions across several periods, which would even further improve efficiency, but it seems the most obvious choice to coordinate on the projects within a period as the horizon for the budget.

where reduced to 20p (10p) in order to compensate for the higher number of projects and leave hourly wages comparable to the other treatments.

Table 8 relates the honesty rates and expected efficiency in the two control treatments to those in the other treatments.

treatment	H_w	H_s	Exp. Efficiency
RAN	7.0%	97.2%	52.4%
EXO	85.7% (88.7%)	84.9% (89.3%)	87.2% (89.1%)
FIX	11.6%	97.9%	55.0%
RLK	12.8%	98.2%	55.9%
CMP	30.9%	93.8%	62.1%
FXI	5.6%	98.8%	52.4%
F4P	19.1%	92.4%	56.6%

Table 8: Share of truthfully represented weak and strong preferences and expected efficiency for FXI and F4P.

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. As can be seen from Table 8, however, there is essentially no effect of such ex–post information. 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 followed a grim trigger strategy, reverting to constantly stating strong preferences once 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. Subjects seem to be impatient and unforgiving. In most treatments, some subjects overestimate their own honesty and underestimate the honesty of the others. They then punish people for being dishonest that are not less dishonest 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.

Concerning reason 2, 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 easy 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 still linked across periods, this would again be difficult and hence we would expect short–term budgeting.

As we can see again from Table 8, there is some increase in H_w compared to RAN or FIX, but the effect is weaker than in CMP. Indeed, the only significant difference is the comparison of H_w with FXI, but H_w is not significantly different from CMP. There is also a notable decrease in H_s, which is, however, still significantly larger than in EXO, though only at $p < 10\%$. Since the increase of H_w is partly cancelled by the decrease in H_s, the increase in efficiency compared to FIX is very small and insignificant.[19]

Therefore, it appears that substantially reducing the coordination problem alone does not enable pairs of players to coordinate on relatively truthful representation of preferences. Compared to the exogenous budgets, however, reducing the coordination problem comes at a price in F4P. The design suggests linking four projects and this implies lower efficiency gains than linking all 40 projects as in EXO. A reasonable benchmark to compare F4P to would hence be a treatment where in each period subjects are restricted to a 2–strong–2–weak–budget. Given the draw of preferences observed in the experiment, if subjects had been under such a rule and had played the Bayesian Equilibrium strategy to state preferences as truthfully as possible, we would have observed honesty rates of $H_w = 81.0\%$

[19]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.

and $H_s = 79.8\%$. So even compared to this benchmark, our subjects exaggerate their preferences quite frequently, but state strong preferences more honestly since they are less constrained in this respect.

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 the third, and partly also the second of the above problems. 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 high preference.[20]

The competition treatment also partly solves the second problem. If a pair forms endogenously, they share two projects each period. Thus they can link these two projects and hence allow each other to state only one strong preference per period. This is, of course, far less effective than linking a larger number of projects across periods. In particular, it forces players to state occasionally a weak preference when their true preference is strong. This explains the lower H_s in CMP.

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 or her. 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.[21]

[20]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.

[21]One subject stated in the questionnaire, that she followed the strategy to state one–weak–one–strong

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 (2005) 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 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, there is 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.[22] 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

and if her partner deviated from this rule, she would switch to another partner for one period. The data shows that she indeed did.

[22] And this implies that players can more often exaggerate their preferences, but also more often have to downplay their preferences.

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 even 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.[23] While the main result of 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 eliminating it 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.

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

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 (2003). An Experimental Study on Storable Votes. NBER Working Paper No. 9982.

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.

Fischbacher, Urs (1999). Z-Tree, Zurich Toolbox for Readymade Economic Experiments, *Working paper Nr. 21*, Institute for Empirical Research in Economics, University of Zurich.

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 (2004). A First Experiment on Qualitative Voting. Mimeo: LSE.

Jackson, Matthew and Hugo Sonnenschein (2005). Overcoming Incentive Constraints. Econometrica, forthcoming.

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

Tyran, Jean–Robert, Huck, Steffen, and Ruchala, Gabriele (2005). Trust, Reputation, and Competition. Mimeo: University of Copenhagen.

A Payoffs from being honest, tough, and nice

An agent's expected payoff of being honest if the other is being 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

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