# Do binding agreements solve the social dilemma ?

Emmanuel Sol, Sylvie Thoron, Marc Willinger

GREQAM

Groupement de Recherche en Economie
Quantitative d'Aix-Marseille - UMR-CNRS 6579
Ecole des Hautes Etudes en Sciences Sociales
Universités d'Aix-Marseille II et III

**Document de Travail**
**n°2006-47**

# DO BINDING AGREEMENTS SOLVE THE SOCIAL DILEMMA ?

**Emmanuel SOL**
**Sylvie THORON**
**Marc WILLINGER**

**December 2006**

# Do binding agreements solve the social dilemma?[1]

Emmanuel Sol[a], Sylvie Thoron[2b], Marc Willinger[c]

December 2006

***Abstract :***

We investigate whether "binding agreements" can provide a solution to the social dilemma that arises in the presence of pure public goods. By signing a binding agreement, players can prevent free riding by the contributors to a public good. However, a well known theoretical result is that the outcome of the endogenous formation of agreements is not necessarily efficient. In our setting, the individual level of contribution to the public good increases with the size of the coalition reaching an agreement and the global coalition is always the socially optimal structure. Agreements form sequentially and the equilibrium outcome is an asymmetric structure, which consists of two coalitions. Our experiment therefore lends force to the theoretical result that outcomes may be inefficient. In fact, we observe an outcome which is even less efficient than that predicted by the equilibrium agreement structure. However, it seems that when subjects reach agreements they do so with the intention of cooperating rather than free riding. Furthermore, it seems that they "learn to cooperate" over time and reach the global agreement more often towards the end of sessions.

Keywords Agreement. Public good. Experiment. Externalities. Sequential game.

JEL Classification C72, C78, C91, C92

a LAMETA, Université de Montpellier, France, sol@lameta.univ-montp1.fr

bGREQAM, Université de Toulon, France, thoron@univmed.fr

c LAMETA, Université de Montpellier, France, willinger@lameta.univ-montp1.fr

[1] We are grateful to the participants of the PLESS seminar in Princeton and of the Interactive workshop in Marseille. This research is in part supported by the funds of the ACI Complex Systems attributed by the CNRS.
[2] While this paper has been written, the author was a visitor at the Institute for Advanced Study in Princeton and at the Department of Economics at Princeton University. She wants to thank the Department for its hospitality and the Institute for its open mind atmosphere.

## 1. Introduction

A standard result in economic theory is that public goods will generally be provided at a level which is lower than, and Pareto inferior to, the socially optimum amount. In the absence of well-designed incentives, agents try to free ride on the contribution of others. If this behaviour is widespread in a community of agents, voluntary contributions will generally be insufficient to produce the optimal amount of public goods. However it has been shown in experiments that the outcome may not be so poor as that predicted by theory.

Experiments involving voluntary contributions to a public good have produced robust stylised facts about the average level of contribution and its evolution with repetition. On average subjects contribute a larger share of their endowment to the public good than predicted under the assumptions of rational and selfish behaviour. The stylised facts that emerge from the numerous experiments for the standard type of environment [3] is that the average contribution is about half the endowment in the first period, and declines with the repetition of the game towards the Nash equilibrium level of contribution, without, however, reaching this level (see Ledyard (1995) and Laury & Holt for a survey of this literature).

While cooperation appears to be prevalent in the early stages of the game, many subjects seem to become less cooperative over time and even end up free riding completely on the contributions of the others. Individual contribution behaviour appears therefore to be largely determined by the observation of past contributions by other group members as well as their own past contributions. Indeed, a large fraction of subjects appear to contribute conditionally (Keser & van Winden, 2000), and few subjects only act unconditionally without adpating their behaviour over time, remaining either cooperative the whole time or always acting as a free rider. In recent years research has focused on ways to improve or sustain cooperation in public goods games. Fehr & Gächter (2000) made an important step by showing that, if free riding behaviour can be punished, there is a marked increase in the average level of contribution. Therefore the threat of individual sanctions, may create sufficient discipline to remove the temptation for agents to free ride.

[3] Yet, most of the available experimental evidence has been obtained for quite particular environments, for which the amount of public good is the output of a linear production technology, i.e. exhibiting constant marginal returns. In particular, public goods games with a unique dominant strategy equilibrium, assume that the amount of public good is a linear function of the total contribution of the group.

While this opens an important research avenue, in practice individual contributions are often unobservable. For example, many donators to charities prefer to remain anonymous, a fact that is well accepted by the community. More important is the fact that complete observability of individual contributions may be too costly to be implemented. For example tax evasion cannot be eliminated completely since costly inspections are limited by the authorities' budget constraint. Even in situations where individual contributions are observable at a reasonable cost, punishment is not necessarily feasible. For these reasons it is worth exploring other tools that can improve the level of cooperation and contribution to the public good within a community of agents.

In this paper we investigate a new mechanism that can lead to higher contribution levels in an experimental game of contribution to a pure public good: binding agreements. In contrast to voluntary contributions, under a binding agreement players have to make a commitment to a fixed level of contribution. The commitment can be made by several players who decide to sign an agreement. International environmental treaties constitute an obvious example. However, in the case of the negotiation on global warming, there is an ongoing debate. On the one hand there are those countries which favour voluntary contributions and these include the United States and certain developing countries like China and, on the other hand there are those such as the European Union, in favour of a binding agreement which specifies targets and time table.

In the theoretical literature, the old debate about the necessity and the efficiency of binding agreements has been revived in the last 20 years. The problem of public goods and externalities has been analyzed by economists in the framework of Samuelson's public good theory (1954) or as developments of the Coase theorem (1961). The Coase Theorem explains how partners can reach an efficient outcome by bargaining, when they can sign a binding agreement at the end. The new theory of coalition formation explained how the existence of these externalities or spillovers could lead to inefficient outcomes. Even when the players have the possibility to sign a global agreement, they may fail to reach this efficient outcome and sign sub-global agreements. This is the problem of participation in agreements.

The numerous contributions to this literature converge towards the same result, but differ in the form of games used to model coalition or agreement formation. One part of this literature uses normal form games which are reduced form models, useful when the objective is not the analysis of the emergence of agreements but the analysis of their stability. The first ancestral exclusive membership game was proposed by Von Neumann and Morgenstern (1944) in their seminal book which marked the beginning of game theory. Later, Hart and Kurz (1983) proposed a variation of the same model. At the same time, D'Aspremont et al. (1983) proposed a simpler game with an open-membership rule, which opened the way for numerous applications to industrial organization (the overabundant theory of stable cartels[4]) and environmental economics (Barrett (1994), Carraro and Siniscalco (1993), Hoel and Schneider (1997),…). A parallel literature uses extensive form games which allow one to describe, some aspects of the bargaining leading to the agreement. See for example Bloch (1996), or Ray and Vohra (1999), (2001). Finally, another part of the literature deals specifically with the complexity of farsightedness in a sequential framework. See for example Chew (1994) or Xue (1998).

In this paper, we adopt the extensive form game approach. We rely on the model of Ray and Vohra (2001) in which the players can reach agreement by making proposals and counter-proposals sequentially. At each period of time, one player can propose an agreement involving each of the players who would be required to sign. If the proposed agreement is accepted by all potential members the agreement is definitely created. If one of the potential members rejects the offer, he has to make a new proposal. The process continues until each agent belongs to an agreement. Then, the agents have to contribute to a pure public good and those who have reached an agreement contribute the amount which maximizes the sum of their partners' and their own payoffs, given what is decided by the other groups. When the game is symmetric, this leads to situations in which all agents who accept to sign an agreement have to contribute the same amount to the public good. This amount increases with the number of agents who sign an agreement. Individual contributions are therefore more costly in large agreements and the total amount of public good provided depends on the structure of agreements. For this reason the social dilemma is not completely eliminated since there is an incentive for small agreements to free ride on larger ones. Since the individual level of contribution is increasing with the size of the agreement, there is an incentive for players to stay alone, or to join few partners in order to free ride on existing bigger agreements. On the other hand, since the amount of public good depends on the size distribution of the agreements, there is an incentive to build up larger groups. There are two

[4] See for example Donsimoni (1985), Donsimoni, Economides and Polemarchakis (1986), Thoron (1998).

extreme structures which could emerge: the global agreement and the set of singletons. In this setting, the global agreement corresponds to a social optimum in the sense that it leads always to the most efficient outcome. In contrast, the set of singletons leads to the worst social outcome. However, Ray and Vohra show that, at the equilibrium of their game, several sub-global agreements can co-exist. Given the parameters we have chosen, the equilibrium structure of agreements is an asymmetric structure with two agreements, in which the smaller group free rides on the larger one. Since the equilibrium structure of agreements differs from the set of singletons, binding agreements increase the level of cooperation in the population.

We present the result of an experiment based on Ray and Vohra's (2001) game. Our experiment compares two protocols of agreement formation: the veto treatment and the dictator treatment, which are equivalent from the theoretical point of view. In the veto treatment, which corresponds to the agreement formation game described above, those individuals who are proposed as members of an agreement can reject the proposal. In contrast, in the dictator treatment, they are obliged to become members of the proposed agreement. The equilibrium prediction is the same for the two protocols. With our choice of parameters there is a unique asymmetric equilibrium agreement in which a small coalition free rides on a large coalition.

We found strong evidence of both cooperative and free riding behaviour. Our experiment supports the theoretical result that the formation of binding agreements leads to inefficient outcomes. We obtain an even worse outcome than that predicted by the equilibrium agreement structure. Indeed, our data shows that the equilibrium outcome is almost never observed. In contrast, the analysis of proposals in sub-games shows that most proposals correspond either to the global agreement or to a structure that incorporates many singletons. It seems that the subjects do not reach agreements to free ride but only to cooperate. Furthermore, they "learn to cooperate" over time and reach the global agreement more often towards the end of sessions.

The rest of the paper is structured as follows. Section 2 presents the extensive game of coalition formation under positive externalities. The game is solved assuming the rationality of players adopted in Ray and Vohra's paper. Section 3 introduces the experimental design. Section 4 presents our main findings about the realized agreements (4.1) and the proposed agreements (4.2), which differ only for the veto treatment. Section 5 is a discussion of different interpretations of the experimental results. Section 6 concludes.

## 2. Theoretical background

### *2-1 The sequential game of agreement formation*

We consider a two-stage model of sequential formation of binding agreements. In a first stage, $n$ identical players have the possibility to sign binding agreements. The outcome is a partition of the set of players. In the second stage of the game, the players voluntary contribute to a public good. The players who have signed the same agreement choose the contributions which maximize the sum of the partners' utilities, given the decision taken by the signatories of the other agreements.

The first stage of the agreement formation game is drawn from Bloch (1995), (1996) and Ray and Vohra (1999), (2001). They propose a sequential game of agreement formation based on a bargaining game à la Rubinstein (1986) and Stähl (1972). Let $N$ denotes a set of players. A protocol designs the order in which the players in $N$ enter into the game to make a proposal or to give an answer to somebody else's proposal. A proposal by player $i \in N$ is an agreement $S \subset N$ to which she belongs and which is characterized by the members' names and a sharing of the agreement's payoff. The agreement can only be signed if the different members have sequentially accepted the proposal. If one member refuses, she has to make another proposal. If they have all accepted, the agreement is signed and the game continues with the players who have not reached an agreement yet. Once an agreement is signed its membership cannot be modified; in other words, there is no renegotiation. The outcome is an agreement structure $\pi = (S_1,\ldots,S_m)$, or, in other words, a partition of the set of players, which means that each player signs one and only one agreement. Formally,

$$\forall k, l = 1,\ldots,m,\ k \neq l,\ S_k \cap S_l = \emptyset \text{ and } \bigcup_{k=1}^{m} S_k\ .$$

This game is an extensive form game whose sub-games start each time a player has to make a proposal. Therefore, for each partition of each sub-set of players $T \subset N$, denoted by $B(T)$, and for each player $i \in N \backslash T$ designated by the protocol to make a proposal, we can define a sub-game denoted by $G^i_{B(T)}$. Players in $N\backslash T$, player $i$ and his potential partners, have not reached an agreement yet and are the "active" players in this sub-game. If all the players in $N\backslash T$ reach an agreement, we will say that the *largest agreement* is reached in the sub-game $G^i_{B(T)}$. If the largest agreement is reached at the beginning of the game, that is in a sub-game $G^i_{B(\emptyset)}$, we will say that the *global agreement* is reached.

When the game is symmetric and the players all identical, the net benefit generated by a given agreement $S_k \subset N$, in a given agreement structure $\pi = (S_1,\ldots,S_m)$, only depends on the number of signatories of each agreement, element of the structure. Bloch (1996) proved that, in that case, the sequential game of agreement formation is equivalent to a simpler game in which each player designated by the protocol chooses a number of partners and the agreement is immediately signed. Indeed, the interests of the player who makes a proposal coincide with those of the partners she chooses and who are in the same position. The outcome of the game is then an ordered sequence of *agreement sizes* which sum to $n$, the total number of players. Ray and Vohra (1999) proved that the endogenous sharing rule in each agreement is then the equalitarian sharing of the net benefit generated among the signatories.

### *2.2 Positive externalities*

Payoffs are generated in the second stage of the game. Ray and Vohra (2001) refer to a model of pollution control. Think for example that a set of partners have signed a binding agreement to improve the quality of air. In order to decrease pollution, they have to decrease their production or to adopt new technologies and this is costly. Let z denote the public benefit of control activity pursued by any particular partner. Let $c(z)$ be the private cost generated by this control activity. We assume that this cost function is quadratic: $c(z) = (1/2)z^2$. The payoff to partner $i$ is then:

$$u_i(z_1,\ldots,z_n) = \sum_{j=1}^{n} z_j - c(z_i)$$

For a given partition of the set of individual players $N$ into $m$ binding agreements $\pi = (S_1,\ldots,S_m)$, the signatories of each agreement $S_i$ of size $s_i$ decide the amount of pollution control each member has to produce $z_i$, given the effort of the other agreements' signatories:

$$\max_{z_i} s_i \left( s_i z_i - c(z_i) + \sum_{\substack{j=1 \\ j \neq i}}^{m} s_j z_j \right)$$

This is a non-cooperative game in which the new players are now the *m* binding agreements. At the Nash equilibrium of this game, each partner of an agreement of size $s_i$ produces an amount $s_i$ of pollution control and enjoys a payoff:

$$u_i(s_1,\ldots,s_m) = \sum_{j=1}^{m} s_j^2 - \frac{1}{2} s_i^2$$

This payoff function generates positive externalities. This means two things. First, for a given structure of agreements, the signatories of small agreements get more than the signatories of large agreements. Secondly, for a given agreement, its signatories are better off if other agreements “merge”. As a consequence, for a given player, the best situation is the agreement structure in which she does not sign any agreement while all the other players sign a unique agreement of size $n - 1$. When the whole set of individual players in *N* sign the global agreement, the different partners share equally the collective optimum, which is $n^2/2$. However, in this situation, each partner would prefer not to sign the agreement in order to free ride on the other signatories. Two opposite forces are at stake. Individual players have an incentive to sign agreements with many signatories in order to efficiently control pollution. At the same time, each individual player has an incentive to free ride and would prefer the other players to sign the agreement without her participation.

### *2.3 Farsighted players*

If the players are farsighted, as it is assumed in Ray and Vohra (2001), a unique Nash equilibrium of the sequential model of agreement formation is determined by backward induction. Let us again consider the symmetrical case. Ray and Vohra show that the global agreement is not necessarily reached. Furthermore, if several sub-global agreements emerge, they appear in size-increasing order. Consider for example, an original set of seven players. At the unique equilibrium, an agreement between two players is signed first, followed by the signature of another agreement by five other players. The smaller group "free rides" therefore on the contributions of the larger group.

Let us see how this happens. Players are farsighted and the game is solved by backward induction. The simplicity of the following argument is a consequence of the linearity of the payoff function. Each decision about the size of the agreement to sign does not depend on the agreements already signed, but depends only on the number of players who did not take a decision yet.

Consider a sub-game $G^i_{B(T)}$ such that a structure of agreements *B* has already been reached. Then, we denote by g(B) the public good contributed by the sub-set of players *T* who have signed one of the *B*-agreements. If there is only one player remaining, $N\backslash T = \{i\}$, of course she forms a singleton and gets: $u_i(B,1) = g(B) + 1/2$. If there are two players left, $N\backslash T = \{i,j\}$, the first player required to take a decision compares two possibilities:

1) Staying alone, in which case her payoff will be $u_i\ (B, 1,1) = g(B) + 3/2$.
2) Signing an agreement with another player to get $u_i\ (B, 2) = g(B) + 2$.

As a consequence, she signs an agreement with the other player who accepts because he gets more tand the stand alone payoff.
If there are three players remaining, the first player who needs to take a decision compares three possibilities:

1) Staying alone and anticipating that the two other players will sign an agreement, in which case her payoff will be $u_i\ (B, 1,2) = f(B) + 4{,}5$.
2) Signing an agreement with another player, $u_i\ (B, 2, 1) = f(B) + 3$
3) Signing an agreement with the two other players, $u_i\ (B, 3) = f(B) + 4{,}5$

and she proposes an agreement to the two other players (assuming that the larger agreement will be chosen whenever there is payoff equality between several agreement structures). The same reasoning applies until the first player who is perfectly farsighted and anticipates the decisions of the following players as previously described. If $n$ = 7, the equilibrium outcome is an agreement structure $B^*$ = (2, 5).

In fact, this equilibrium depends on three assumptions about the players' behaviour. The players' farsightedness allows them to anticipate all the subsequent actions of the other players. However, this is not sufficient to apply backward induction. A second necessary assumption is that all the players are selfish maximizers, i.e. individual payoff maximisers. The last assumption is that players' rationality is common knowledge. We will denote these three assumptions about the players' rationality as follows:

Assumption A1: Selfishness

Assumption A2: Farsightedness

Assumption A3: Homogeneity

We will refer to individuals having the rationality generated by these three assumptions as having R&V's behaviour and being R&V's players.

## 3. Experimental design

Our experiment compares two theoretically equivalent protocols for agreement formation: the veto treatment and the dictator treatment. In the veto treatment subjects selected to become a partner of a proposed agreement can reject the offer, in which case a new proposal is made by another subject. In the dictator treatment subjects that are selected to become a partner of a proposed agreement cannot refuse. Therefore, in the dictator treatment the number of proposals in a round is always equal to the number of final coalitions, whereas in the veto treatment the number of observed proposals is generally larger than the number of final coalitions of the realized structure of agreements. We therefore need to analyse the veto treatment, to study rejected offers. separately Each independent group of subjects played the agreement game for several rounds to allow for learning. The number of repetitions in each independent group was random. The subjects knew that an additional round would be played with some probability. As a result, the number of rounds differs across independent groups. In addition to our main treatment variable "veto versus dictator", we compare for each treatment fixed groups with randomly matched groups. In fixed groups the group of subjects remains the same in all rounds. In randomly matched groups, subjects are assigned randomly to a new group of players after each round. We thought that fixed groups could behave differently, because subjects may react to previously observed agreements in their group. For example, a subject who is disappointed by previous agreements could be inclined to act in a more selfish way in future rounds, if he is selected to make a proposal. This effect can even become stronger in the dictator treatment. We collected 8 independent groups of subjects for the veto treatment (6 fixed, 2 random) and 7 independent groups for the dictator treatment (5 fixed, 2 random). In each session, 14 participants played the agreement game, and were divided into groups of seven subjects. Table 1 summarizes the experimental design. Note that the number of rounds in sessions for random groups is larger because we counted this type of session as a single independent observation. The number of rounds of course is equal to the number of observed final agreements, except for random groups for which the number of rounds must be multiplied by 2. Table 1 indicates the number of observed proposals in each group. Of course the number of proposals is smaller for groups in the dictator treatment, since proposals cannot be rejected.

[INSERT TABLE 1]

The experiment was carried out using standard procedures. We developed a specific computer program for running the experiment. Our experimental procedure follows closely the protocol described in Ray and Vohra's model. In each period, one of the subjects who does not yet belong to an agreement is randomly selected to make a new proposal. A proposal consists simply in stating an agreement size. The members assigned to the agreement are chosen randomly among the players who do not yet belong to an agreement, and are privately informed that they have been selected.

In the veto treatment subjects must state privately if they agree or not to sign the agreement. If all of them accept the proposal, the agreement is signed. On the other hand, if one of the assigned players rejects the offer, he will have to submit a new proposal. If more than one player rejects the offer, one of them will be selected randomly to make a new proposal. Our procedure differs slightly from the theoretical procedure, where the proposed members are asked sequentially whether or not they accept to become a member of the proposed agreement. Under the sequential protocol the first player who rejects the offer is the one that makes a new proposal. To gather more data about subjects' reactions to proposed agreements, we chose instead to ask all the proposed members to answer simultaneously. In case of a rejection, subjects only knew that the proposed agreement was rejected without knowing by how many subjects[5]. Does the difference between the experimental design and the theory affect the subjects' incentives in the game? It may be the case that a subject's incentive to reject be affected by the fact that he cannot be sure that he will be the one who makes the next proposal. However, whether this uncertainty increases or decreases the incentive to reject is unclear.

[5] Of course a subject who did reject but was not selected to make a new proposal could easily infer that there was at least one other subject who rejected. Note that this is also the case in the model.

For simplicity let us call "player 1" the subject who makes the first proposal, "player 2" the subject who makes the second proposal, and so on. Subjects were told at the beginning that they would be randomly assigned to a group of 7 subjects. They could rely on a table showing their payoff according to their group size and the final agreement structure (see table 2). Each period started by allocating randomly to each subject a "name" chosen from a list of seven names {A, B, C, D, E, F, G}. Then, one of the names was selected randomly to be "player 1", i.e. the subject who proposes the first group. The decision task was simply to choose an integer number between 1 and 7. If the chosen number was equal to one, a singleton was formed. If the chosen number was larger than one, the potential partners of the agreement were randomly selected among the 6 remaining subjects. Once subjects were selected for agreement 1, in the veto treatment they were asked individually whether they accepted or rejected this proposal. If all the proposed subjects accepted, the first agreement came into force . However,, if one or more subjects rejected the offer, the group was not created. Then one of the subjects who decided to reject was selected (randomly) to make a new proposal. This process was repeated until each subject was partner to an agreement or a singleton. Once all subjects were assigned to an agreement, the period ended and the gains were announced to the subjects on their computer screen. The gains were displayed according to agreement size.

[INSERT TABLE 2]

## 4. Results

### *4.1 Realized agreements*

Figure 1 compares the cumulative frequency distributions of accepted agreements in all groups for the random and the fixed group conditions. The null hypothesis of equal frequency distributions under the random matching condition and the fixed group condition cannot be rejected (Kolmogorov-Smirnov, two-sided, 10%). We decided therefore to pool the data from the fixed and the random condition for each treatment.

[INSERT FIGURE 1]

*Result 1* : The most frequently realized "agreement" is the singleton.

Figure 2 displays the frequency distribution of accepted agreements for all groups, for the veto (2a) and the dictator (2b) treatments. In the veto treatment more than 70% of the realized agreements are singletons, compared to 35% for the dictator treatment. However, we cannot reject the null hypothesis of equal size distributions (Kolmogorov-Smirnov, two-sided, 1%). Since there are many agreement structures that contain at least one singleton (11 out of 15 possibilities), singletons are more likely to emerge if subjects had decided on a random basis. Furthermore, if subjects have a tendency to free ride, they will be tempted to propose small agreements where they contribute less. We need therefore to take a closer look at the frequencies of realized agreement structures.

[INSERT FIGURE 2]

*Result 2 :* We observe wide variations in the realized agreements structures. The equilibrium structure (2,5) is observed 11 times (9.6%) in the dictator treatment and never in the veto treatment. The modal structure is the global agreement in both treatments. In the veto treatment more than 50% of the agreement structures contain three or more singletons, compared to less than 8% in the dictator treatment.

[INSERT FIGURE 3]

Figure 3 shows the frequency distribution of agreement structures. Few structures (4.5% for the veto treatment, 25% for the dictator treatment) involve only two agreements, in contrast to the theoretical prediction. The total payoff of the structure has no significant effect on its frequency of appearance (see result 4). Furthermore, low payoff structures are quite frequent for the veto treatment.

*Result 3* : The average gains achieved are higher than those predicted in the dictator treatment and lower than predicted in the veto treatment.

The average performance is equal to 108.19 and 145.17 for the veto and the dictator treatment, respectively, a significant difference (Mann-Whitney, one-sided, 1%). These numbers correspond to 79.49% and 105.96% respectively of the predicted performance (equal to 137, which corresponds to the performance at the equilibrium (2,5)). Only one of the veto groups reaches a better performance than the equilibrium level, and only one of the dictatorial groups has a lower performance than the equilibrium prediction. The performance is significantly lower than equilibrium performance for the veto treatment (Binomial test, one-sided, 3.5%) and significantly larger than equilibrium performance for the dictator treatment (Binomial test, one-sided, 6.2%). We conclude that binding agreements fail to solve the social dilemma in the veto treatment. In contrast, in the dictator treatment, the gap between the social optimum performance and the equilibrium performance is reduced by binding agreements.

*Result 4*

Over the 3 last periods the frequency of the global agreement increases in both treatments.

In table 3, we aggregate all agreement structures that contain at least three singletons. We compare the frequencies of these agreement structures with the frequency of global agreements and the remaining agreement structures (others). The table compares all periods with respect to the three last periods. The frequency of the global agreement increases for both treatments in the last three periods (Wilcoxon signed rank test, $p = 0.0547$ for the veto treatment, $p = 0.1094$ for the dictator treatment). For the veto treatment we also observe that the frequency of agreement structures containing at least 3 singletons decreases.

[INSERT TABLE 3]

*Result 5* : Agreement structures with low payoff disparity among members are more likely to emerge.

Table 4 shows the Gini index associated with each agreement structure and the corresponding observed frequencies for each treatment. For the veto treatment the frequencies are independent of the corresponding total payoff, in contrast to the dictator treatment. Note that for the veto treatment, extreme agreement structures for which the Gini index is close to zero, have a high frequency.

[INSERT TABLE 4]

We regress the frequency of observed agreement structures, on the corresponding Gini coefficient and total payoff of the structure. We include both the Gini index and the squared Gini index, to account for the fact that the relation is not necessarily linear. The results are summarized in table 5. For both treatments the Gini coefficient is negatively related to the frequency of realized agreements (ranked according to payoff disparity). Furthermore, the Gini coefficient has a significant second order effect, which implies that the frequency of agreement structures increases when payoff disparity becomes very high. As can be seen from table 4, the two agreement structures with the highest Gini index ((1,1,5) and (1,1,1,4)) have a relatively high frequency. From table 5, we see that total payoff is highly significant for the dictator treatment, but not for the veto treatment.

[INSERT TABLE 5]

### *4.2 Proposed agreements*

Our preliminary analysis of realized agreements does not take into account the fact that, in the veto treatment, many proposals were rejected. Actually, large agreements were frequently proposed and rejected, revealing the conflict between individual and group interest. While the realized agreements get clearly closer to the social optimum under the dictator treatment, the cooperative intentions are not necessarily weaker in the veto treatment. We analyse these intentions by looking at initial proposals at the beginning of each round. In Section 2 we pointed out that initial proposals are particularly interesting since they tend to reveal that the type of behaviours we described are consistent with three different proposals. For later proposals, which are influenced by the sequence of agreement formation, we compare the proposed agreement sizes to the predictions of Section 2.

#### *4.2.1 Initial proposals*

*Result 6*: In both treatments, the global agreement is the most frequently proposed agreement at the beginning of rounds, followed by the singleton.

We assume that initial proposals in each round provide an indication of subjects' intentions to cooperate or to act strategically. Clearly the cooperative intention is the strongest. However since singletons are often formed, result 6 illustrates the conflict between collective and individual interests. Figure 4 shows the frequency distribution of proposals ranked from the largest to the smallest agreement, at the beginning of each round. Clearly, the singleton (the smallest "agreement") and the global agreement are the most frequently proposed agreements. For the veto treatment, 57% of the initial proposals are global agreements, and 25% are singletons. The corresponding frequencies for the dictator treatment are 45% and 32% respectively. Other agreement sizes have a very low frequency, except agreements of size 2 which was proposed in the 12% of the cases in the dictator treatment.

INSERT FIGURE 4

Although we decided to pool the data from the fixed and the randomly matched groups, it is worth mentioning that fixed groups behave more cooperatively than random groups, as can be seen from table 6. Global agreements are more frequent and singletons are less frequent in fixed groups than in random groups. We cannot test for significance however, since we have only two independent observations in the random treatments. Table 6 and figure 4 also show a difference in the cooperative intentions between the veto and the dictator treatments. Global agreements are more frequently proposed in the veto treatment, and singletons are less frequently proposed. We reject the null hypothesis of equal frequencies of global agreement proposals in initial rounds for the veto and the dictator treatments (Wilcoxon Mann-Whitney, two-sided, 5%), and of equal frequencies of singletons (Wilcoxon Mann-Whitney, two-sided, 10%).

INSERT TABLE 6

In the veto treatment 73.80% of the (non-singleton) initial proposals were rejected. Since agreement sizes between 2 and 6 were very rarely proposed at the beginning of rounds, rejection rates across agreement sizes cannot be compared. Furthermore larger agreements are more likely to be rejected, even if only a few subjects act myopically. Note that global agreements as initial proposals are less frequently rejected by fixed groups (62% rejected) than by random groups (81% rejected).

Since the global agreement was the most frequently proposed and rejected agreement, of course a rejection was followed most of the time by a proposal of a smaller agreement. The rejection of the global agreement was frequently followed by the formation of a singleton[6] (more than 50% of the cases overall).

*4.2.2 Proposals in sub-games*

Table 7 gives the proportion of proposals for various sub-games. Each sub-game is characterized by the number of subjects who are not yet partners of an agreement (row numbers). For instance the row labelled 5 corresponds to situations in which the first two subjects have already signed a binding agreement, either an agreement of 2 subjects or two singletons. The possible agreement sizes in this example are 5, 4, 3 2 and 1. The last column of the table gives the percentage of "extreme" proposals, i.e. the largest possible agreement in the sub-game or a singleton. Obviously for each sub-game, the most frequently observed proposal is either the largest agreement of the remaining players or the singleton. Together, these two possible proposals represent at least 70% of all proposals in each sub-game and in each treatment.

INSERT TABLE 7

Table 8 shows the aggregated frequency of extreme proposals (singleton + largest agreement) for each sub-game and each independent subjects' group. Table 9 (appendix) provides detailed frequencies about singletons and global agreements for each sub-game size and independent player group. Except for sub-game of size 3, the frequencies of extreme proposals are the same in both treatments (Wilcoxon Mann-Whitney, 5%). In sub-games of size 3 only extreme agreements were proposed in the dictator treatment, while intermediate coalitions of size 2 emerged sometimes in most groups in the veto treatment.

INSERT TABLE 8

[6] In a few cases the rejection of the global agreement was followed by the same proposal.

Table 8 clearly shows that in all independent groups and for all sub-game sizes there is a strong tendency to propose extreme agreements, as predicted by Proposition 2. In order to test for significance of the observed tendency we use Selten's (1991) measure of predictive success, $S = h - a$, where $h$ is defined as the *hit rate* (frequency of compatible proposals) and $a$ as the *area* (frequency of random compatible proposals). The *area* measures the possibility that extreme proposals could have been chosen randomly. For example, for sub-games of size 5, if we assume that a subject who chooses randomly proposes each possible size with equal probability, there is a probability of 2/5 that he proposes a compatible agreement at random. Therefore we need to correct the hit rates associated with each sub-game size by the frequency of randomly chosen sizes. Of course $a$ gets larger as sub-games get smaller, leading to $S = 0$ for sub-games of size 2. All measures of predictive success are significantly positive (Binomial test, one-sided, 5%) for each sub-game size of each treatment. We conclude therefore that extreme agreements are not proposed at random, but are intentionally proposed by subjects.

## 5 Discussion

### *5.1 Cooperation and free riding*

The results described in the previous section constitute strong evidence of the co-existence and combining of cooperative behaviour and an incentive to free ride. Indeed, we observed that the global agreement is the initial proposal in the majority of cases and the most frequent realized agreement. Furthermore, the largest agreement is, in each sub-game, the most frequent or the second most frequent proposal (after the singleton). All of which suggests cooperative behaviour. However, singletons appear with the highest frequency in all agreement structures, indicating that free-riding behaviour seems also to be widespread in our data. The proposal of a global agreement is very often rejected and most of the time this rejection is followed by the formation of a singleton. In each sub-game the singleton is the most frequent or the second most frequent proposal (after the largest agreement).

We suggest that two forces are at work: the incentive to cooperate which drives the subjects to propose an agreement to all remaining potential partners and the incentive to free ride, which drives them to become singletons. The role of these two forces - an incentive to cooperate opposed to an incentive to free ride - was described in Ray and Vohra's paper and in several other papers demonstrating the inefficiency of binding agreements for solving the social dilemma. However in our experiment, the equilibrium predicted by Ray and Vohra, is never observed for the veto treatment and appears with a low frequency in the dictator treatment. The observed outcome for the veto treatment is even less efficient than predicted by the equilibrium. How can we interpret this difference from the theoretical result and this even greater inefficiency?

### *5.2 Other assumptions about the subjects' rationality*

In this section we try to determine which of the previous assumptions characterizing R&V's behaviour, correspond to a bad predictor for our subjects' behaviour?

#### *5.2.1 Cooperative players*

First, let us consider the case where the players are farsighted (A2) and believe in a common rationality (A3) but are not selfish. Since these players want to maximize the sum of the payoffs, we will define them as *cooperative*.

Assumption A1': The objective of a *cooperative player* is to maximize the sum of all players' payoffs.

What would constitute an equilibrium under the foregoing assumptions? Consider a group *T* of players. Whatever the organization of the rest of the population *N**T*, the sum of payoffs is always higher if players in *T* sign a unique agreement than if they sign two distinct agreements, whatever this partition of *T* into two agreements is. Consider that $T = S_i \cup S_j$, this can be verified simply by comparing the sum of payoffs of players in *T* when $S_i$ and $S_j$ are disjoint:

$$s_i u_i(s_i, s_j, \ldots) + s_j u_j(s_i, s_j, \ldots) = (s_i + s_j) g(B(N/T)) + (s_i + s_j)(s_i^2 + s_j^2) - \frac{1}{2}(s_i^3 + s_j^3)$$

and when they are joint:

$$s_i u_i(s_i + s_j, \ldots) + s_j u_j(s_i + s_j, \ldots) = (s_i + s_j) g(B(N/T)) + (s_i + s_j)(s_i + s_j)^2 - \frac{1}{2}(s_i + s_j)^3 .$$

Obviously the latter sum is greater than the first, whatever the sizes $s_i$ and $s_j$ . Using backward induction, this leads to the result that, in each sub-game $G^i_{B(N/T)}$, the first player to make a proposal proposes a unique agreement to the *t* remaining players. If a cooperative player has to respond to a proposal, for the same reason, she will accept to sign an agreement involving all the remaining players and will reject every other proposal in order to propose this agreement herself. Therefore, the following Proposition holds:

*Proposition 1: If all players are cooperative, at equilibrium the largest agreement is proposed and signed in each sub-game and hence, the global agreement is proposed and signed immediately.*

### *5.2.2 Myopic players*

In this sub-section we keep assumptions A1 and A3 but instead of farsightedness (assumption A2 ), we assume that players are myopic. How would myopic players behave in this game? In general, myopic behaviour is understood as to be the behaviour of a player who takes his decision by reacting to the existing situation without anticipating the other players' reactions to his own choice. Given that she anticipates no response, she plays her best response. In our framework, however, myopic behaviour has to be defined more carefully for two reasons. The first reason is that the behaviour will depend on the player's position at the outset. Does she have to make a proposal or to respond to a proposal? But the behaviour depends also on the myopic players' conjectures when they consider the consequences of their choices. Indeed, in our framework, it is not realistic that myopic players have no conjectures at all. They do have conjectures but myopic conjectures. In order to define precisely myopic behaviour, we take into account the player's position in the game and her conjecture.

Myopic behaviour is characterized by the following assumptions A2':

- A player who makes a proposal does not anticipate the possibility of subsequent counter-proposals by other players.
- A player who has to respond to a proposal does not anticipate any other counter-proposal, except her own counter-proposal if she decides to reject.

What would constitute an equilibrium under these assumptions? The first player to make a proposal chooses the coalition size which maximizes her payoff as a member. Therefore, in each sub-game $G^{i}_{B(N/T)}$, her objective is to solve:

$$Max_{s_i} u_i = g(B) + s_i^2 + t - s_i - \frac{1}{2} s_i^2$$

The solution to this program is always to propose the coalition of the *t* remaining players in the sub-game. If a myopic player has to respond to a proposal, for the same reason she will accept a coalition of all the remaining players and will reject every other proposal in order to propose herself that coalition. Obviously, if all the players are myopic, the global agreement is proposed and signed immediately, like in the case where all players are cooperative,

*Proposition 2: If all players are myopic, at the equilibrium the largest agreement is proposed and signed in each sub-game and the global agreement is proposed and signed immediately.*

### *5.2.3 An alternative to homogeneity*

Let us turn now to assumption A3. How is the equilibrium modified when players consider that the population is not homogeneous? For example, consider a selfish and farsighted player. In a homogeneous population, it may be that his equilibrium strategy is to propose a sub-agreement in a given sub-game. In this case, the formation of a singleton is not the best strategy because the remaining players would be induced to sign several agreements. Now, assume that this farsighted player believes that all other players are either myopic or cooperative. Then, she believes that she has the opportunity to stand alone in a singleton while all remaining players will sign an agreement, which corresponds to the most beneficial situation for her. Call this player *a free rider*. The free rider thinks that he is more “clever” than the other players.

Assumption A3’: A free rider is a selfish farsighted player who believes that all the other players are cooperative or myopic.

In each sub-game $G^{i}_{B(N/T)}$, the free rider believes that the remaining players will sign an agreement and she maximizes:

$$Max_{s_i} u_i = g(B) + s_i^2 + (t - s_i)^2 - \frac{1}{2} s_i^2$$

As a result, she refuses every proposal and always proposes the singleton. Then, if *k* players are free riders and *n – k* myopic or cooperative, the following result holds:

*Proposition 3: If k players are free riders and n – k are myopic or cooperative, at the equilibrium the agreement structure is formed by k singletons, which form first, followed by the signature by n – k players of the unique agreement.*

This was the simplest case to consider. We could also have allowed the players' conjectures to evolve. For example, myopic players could become free riders or free riders could learn that there are not isolated. While the characterization of such an equilibrium would be interesting, although much more complex because of incomplete information, it is beyond the scope of this paper to investigate such an issue. All we wanted to show is that if the players' behaviours correspond to one or the other of the three types described previously, the only proposals would be, in each sub-game, the singleton or an agreement signed by all the remaining players. This prediction is close to what we observed in sub-section 4.2.2.

*Proposition 2: If the players are either cooperative, myopic or free riders, in each sub-game each proposal is either the singleton or the largest agreement.*

The following table summarizes, for each type of behaviour, the proposals at the equilibrium in the different sub-games in the case of a population of seven players. Note that, a player who adopts R&V's behaviour is only differentiated by her action in sub-games with 7 players left. Then, the R&V's player proposes an agreement of size 2, while the free rider becomes a singleton (1) and the cooperative or myopic player proposes the global agreement (GA). In sub-games of size 6, a R&V's player acts like a free rider and becomes a singleton. The cooperative or myopic players propose the largest agreement (LA). In sub-games with 5 or less players left, R&V's player acts like a myopic or cooperative player and she proposes the largest agreement of the remaining players. The formation of singletons in these sub-games is only consistent with the free rider rationality.

| | R&V’s behaviour | Myopic or Cooperative behaviour | Free rider |
|---|---|---|---|
| 7 | 2 | LA (GA) | 1 |
| 6 | 1 | LA | 1 |
| 5 and fewer | LA | LA | 1 |

Table 1: Proposals depending on behaviour type and sub-game

Of course, this shows that, in all proper sub-games with less than seven subjects, we cannot distinguish between R&V’s behaviour and other types of behaviours. However, even when the subjects’ proposals coincide with R&V’s equilibrium proposal, they might be generated by a mix of different behaviours. Our data shows that subjects’ initial proposals differ from the proposal predicted by R&V’s behaviour. The global agreement and the singleton appear with a high frequency as first proposals. Furthermore, we observe proposals of the largest agreement in sub-games of size six and singletons formed in sub-games of size five and fewer. Such proposals are compatible with the three alternative types of behaviour that we described : cooperative, myopic and “basic” free riding.

Obviously other sorts of behaviour could coincide with observed proposals. For example, we will discuss below the question of reciprocity. However, the important point here is the difficulty generated by this multiplicity of behaviours. An interpretation of the result above is that the subjects do not play the Nash equilibrium because they have difficulties coordinating. They know that their potential partners do not necessarily behave in the same way. At the beginning, a very careful selfish farsighted subject would try to free ride with a proposal to sign an agreement with only one another partner. However, a cooperative subject would not accept a sub-agreement of two partners if other potential partners are still available. As an argument in favour of this interpretation, in the dictatorial treatment in which this incomplete information is partially solved because the subjects do not have to speculate about the possible counter proposals to their proposals, R&V’s equilibrium was observed, though with a very low frequency.

The main question is to understand how the subjects get round this problem of incomplete information. We try to give an answer to this question in what follows.

### 5.3 Non-maximizing behaviour

#### *5.3.1 Inequity aversion*

The analysis of proposals in sub-games provides strong evidence that subjects focus on two extreme proposals: the coalition of all remaining subjects and the singleton. While we provided several possible behavioural hypotheses that are consistent with the observed patterns, other hypotheses could be suggested. In particular, if subjects are concerned not only by their individual payoff, but also by other players's payoff, they could act in a way that is consistent with our data. If subjects dislike inequalities, as proposed in the seminal models proposed by Fehr and Schmidt (1999) and Bolton and Ockenfels (2001), subjects have an incentive to reach "symmetric" agreement structures. Indeed, an interesting finding in our experiment is that the frequency of appearance of agreement structures is correlated with the Gini index which measures payoff disparities between coalitions of players. Agreement structures with low payoff disparity among members are more likely to emerge.

Another non-maximizing behaviour is often observed in these experiments on voluntary contributions. Subjects use contributions to send signals to their partners or to react to their partners' contributions. Keser & van Winden (2000) say that subjects appear to contribute "conditionally". We explain in what follows that this can play an important role in our experiment as well, but with a different outcome.

#### *5.3.2 Another game*

Another explanation for the fact that our subjects do not play the Nash equilibrium of R&V's game could be that they are not playing the same game. It may be that the subjects "simplify" the game and think just in terms of two actions: "cooperation" or "defection". Indeed, the proposal of an agreement could be interpreted as an intention to cooperate, rather than a tool in order to "free ride efficiently". Subjects who want to free ride would therefore choose a singleton in order to "defect". In this interpretation, each subject chooses between two actions:

1) the subject proposes the largest agreement of the available partners in each sub-game when she wants to “cooperate” or
2) the subject forms the singleton when she wants to “defect”.

In this simplified game, payoffs are given in the following Table:

INSERT Table 11

Note that, for the alternative rationalities Cooperative, Myopic and Free Rider the results of Section 5.2 are unchanged. However under assumptions A1, A2 and A3 characterizing Ray and Vohra’s rationality the equilibrium structure is (1,1,5). In the experiment, (1,1,5) appears with a low frequency (see Figure 3) but we observed in Section 4 that the quasi totality of structures to appear are composed with a unique non trivial agreement and singletons.

We would like to highlight the peculiarities of this simplified game. Consider a subject A who has chosen to be cooperative but whose proposal of the largest agreement has been rejected by a potential partner P. A cannot be forced to be in a less favourable position than subject P who decided to free ride and became a singleton. In the experiment, subject A can always be, at the end of the round, in the same position as the free rider. She can simply decide to become a singleton as well. As a consequence, a cooperative subject could never lose in comparison with a selfish subject, by playing a strategy Tit for Tat. Defection is less frustrating for cooperators. This may explain that the global agreement is the most frequent first proposal (Result ?). Furthermore, in an experiment, this “costlessness” of proposals to cooperate is true till the last round. This may explain that we do not observe any “end-effect”[7]. On the contrary, it is said in Result 4 that the frequency of appearance of the global agreement increases during sessions of our experiment and with it, the average performance. It seems that the subjects “learn to play cooperative”. Many singletons are formed in the first rounds but the frequency of the global agreement increases in the last rounds.

## 6. Conclusion

In this paper we have explored the capacity of binding agreements to mitigate the social dilemma which arises in pure public good provision. In our experiments, the subjects have the

[7] We refer to the well known stylised fact obtained in experiments on voluntary contributions to a public good such that that subjects “learn to play Nash”. They start contributing a lot but then, they adjust their contributions to the equilibrium level.

possibility to sign binding agreements, which guarantee the collective optimum for them and their partners. The global agreement, which corresponds to the collective optimum for the whole group is frequently proposed, especially in the first round, but it is also very frequently rejected in the veto treatment. However, the frequency of realization of the global agreement increases over time. Therefore, on one hand our data provide an optimistic point of view on the capacity of binding agreements to mitigate the social dilemma. On the other hand, we also observe contrary evidence, since the singleton is the second most frequently realized "agreement".

A central result of the literature on endogenous formation of coalition is that binding agreements do not necessarily lead to an efficient outcome. In the model proposed by Ray and Vohra (2001), the players have the possibility to sign binding agreements to contribute to a public good. They show that the incentive to sign large agreements in order to "internalize" the positive externality does not eliminate the incentive to free ride. Their message is that the incentive to free ride is not solved by the signature of binding agreements because it is shifted from the individual level to the group level. In an example with seven players, the equilibrium agreement structure is formed by an agreement of two players and another one of five players. The smaller agreement is signed before the larger one and its partners free ride on the partners of the latter to obtain a larger benefit.

Our experiment lends partial support to this description, but the two driving forces of this process are somehow exacerbated in our data. Indeed, we observe that subjects have an incentive to cooperate by proposing an agreement to all the available partners, but at the same time they also have a strong incentive to free ride and stay as singletons. However, the theoretical equilibrium as described by Ray and Vohra is never observed in the veto treatment and appears with a low frequency in the dictator treatment. As a consequence, the result is even more inefficient in our veto treatment than at the equilibrium predicted by the theory. We give two interpretations for this. The first interpretation is that the subjects have difficulties to coordinate because they know that their potential partners do not have necessarily the same rationality. The other explanation, which is or is not a consequence of the first problem of incomplete information, is that the subjects "simplify" the game and just think in terms of two actions: "cooperation" or "defection", as in a multi-person prisoner dilemma game. Indeed, the analysis of proposals in sub-games provides strong evidence that the subject focus on two extreme choices: the largest agreement of all the available partners or the singleton. The proposal of an agreement is associated with cooperation and it does not seem that the subjects free ride in groups.

The last main result concerns the evolution of the outcome during each session. It seems that the frequency of appearance of the global agreement increases during the session and with it, the average performance. This is in contradiction with the well known stylized fact obtained in experiments on voluntary contributions to a public good which is interpreted as a subjects' ability to "learn to play Nash". They start contributing a lot but then, adjust their contributions to the equilibrium level. In our experiment where the subjects have to sign binding agreements to contribute, it seems that the opposite occurs and that the subjects "learn to play cooperative". Many singletons are formed in the first rounds but the frequency of the global agreement increases in the last rounds. We give an interpretation related to the peculiarity of the game simplified by the subjects.. On the one hand, because the subjects have to make proposals, cooperation is contingent and can be used as a signal to other subjects with no cost till the end of the session. On the other hand, it is so difficult to "free ride efficiently" that defectors can be converted to cooperation.

## References

Barrett, S. (1992), *“Self Enforcing International Agreements”,* Oxford Economic Papers, 46, 878-94.

Bloch F. (1996), *"Sequential Formation of Agreements with Externalities and Fixed Payoff Division",* Games and Economic Behavior, 14, 90-123.

Bloch F. (1997), *"Non-cooperative models of coalition formation in games with spillovers",* 311-352 in New directions in the economic theory of the environment, Edited by C. Carraro and D. Siniscalco, Cambridge University Press.

Bolton, G. E. and Ockenfels, A. (2001), *"A Theory of Equity, Reciprocity and Competition",* American Economic Review.

Carraro, C. and D. Siniscalco (1993), *“Strategies for the International Protection of the Environment”,* Journal of Public Economics, 52, 309-28.

Chwe M. (1994), *“Farsighted Coalitional Stability”,* Journal of Economic Theory, vol 63, 299-325.

D'Aspremont, C., Jean J. Gabszewicz, A. Jacquemin et J.W. Weymark (1983) *"On the Stability of Collusive Price Leadership",* Canadian Journal of Economics, pp. 17-25.

Donsimoni, M.-P. (1985) “Stable Heterogeneous Cartels”, International Journal of Industrial Organization, vol. 3(4), pp. 451-467, December.

Donsimoni, M., N.S., Economides and H. M. Polemarchakis (1986) *“Stable Cartels,”* International Economic Review 27, 317-27

Fehr E., Gächter S.. (2000), *"Cooperation and Punishment in Public Goods Experiments",* American Economic Review.

Fehr E., Schmidt, K. M. (1999), *"A Theory of Fairness, Competition, and Cooperation",* Quarterly Journal of Economics 114, 817-868.

Hart and Kurz, (1983), *"Endogenous Formation of Coalitions",* Econometrica, Vol. 51, No 4, 1047-1064.

Hoel, M. & K. Schneider (1997), *"Incentives to Participate in an International. Environmental Agreement",* Environmental and Resource Economics, 9, 153-170.

Keser C., Van Winden F. (2000), *"Conditional Cooperation and Voluntary Contributions to Public Goods",* Scandinavian Journal of Economics, 102, 23-39.

Laury Susan K. and Charles A. Holt, "Voluntary Provision of Public Goods: Experimental Results with Interior Nash Equilibria," forthcoming in *The Handbook of Experimental Economics Results*, C. Plott and V. Smith, eds.

Ledyard J. (1995), *"Public Goods: A survey of Experimental Research",* in Kagel, J. and Roth, A. (eds.), The Handbook of Experimental Economics, Princeton University Press.

Ray D. and R. Vohra (1997), *"Equilibrium Binding Agreements",* Journal of Economic Theory, 73, 30-78.

Ray D. and R. Vohra (1999), *"A Theory of Endogenous Agreement Structure",* Games and Economic Behavior, 26, 286-336.

Ray D. and R. Vohra (2001), *"Coalitional Power of Public Goods",* Journal of Political Economy, Vol 109, n 6.

Rubinstein, A. (1982), *"Perfect Equilibrium in a Bargaining Model",* Econometrica, 50, 97-109.

Selten, R., (1991), *"Properties of a measure of predictive success",* Mathematical Social Sciences 21, 153–167.

Stähl I., (1972), *Bargaining Theory*, Stockholm School of Economics.

Thoron, S., (1998), *“Formation of a Coalition-Proof Stable Cartel”,* Canadian Journal of Economics, Vol. 31 n°1, 63-76.

Von Neumann, J. and O. Morgenstern (1944), *Theory of Games and Economic Behaviour*, Princeton: Princeton University Press.

Xue L. (1998), *“Coalitional Stability under perfect forsight”,* vol 11, 603-627.

Yi, (1997), *"Stable Coalition Structure with externalities",* Games and Economic Behavior, 20, 201-237.

| Session number | Group | Number of rounds | Number of proposals | Veto / Dictator | Random / Fixed |
|---|---|---|---|---|---|
| 2 | V1 | 13 | 140 | veto | random |
| 3 | V2 | 8 | 49 | veto | fixed |
| 3 | V3 | 8 | 67 | veto | fixed |
| 4 | V4 | 11 | 76 | veto | fixed |
| 4 | V5 | 11 | 65 | veto | fixed |
| 5 | V6 | 14 | 173 | veto | random |
| 6 | V7 | 10 | 48 | veto | fixed |
| 6 | V8 | 10 | 84 | veto | fixed |
| 7 | D9 | 15 | 25 | dictator | fixed |
| 7 | D10 | 15 | 28 | dictator | fixed |
| 8 | D11 | 10 | 48 | dictator | random |
| 9 | D12 | 10 | 40 | dictator | random |
| 10 | D13 | 15 | 33 | dictator | fixed |
| 11 | D14 | 15 | 29 | dictator | fixed |
| 11 | D15 | 15 | 27 | dictator | fixed |

Table 1 : Experimental design

| # | Structure | $s_1$ | $s_2$ | $s_3$ | $s_4$ | $s_5$ | $s_6$ | $s_7$ | Total payoff |
|---|---|---|---|---|---|---|---|---|---|
| 1 | (7) | 24,50 | | | | | | | 172,00 |
| 2 | (1,6) | 36,50 | 19,00 | | | | | | 151,00 |
| | (2,5) | 27,00 | 16,50 | | | | | | 137,00 |
| | (4,3) | 20,50 | 17,00 | | | | | | 130,00 |
| 3 | (1,1,5) | 26,50 | 26,50 | 14,50 | | | | | 126,00 |
| | (1,2,4) | 20,50 | 19,00 | 13,00 | | | | | 110,50 |
| | (1,3,3) | 18,50 | 14,50 | 14,50 | | | | | 106,00 |
| | (2,2,3) | 15,00 | 15,00 | 12,50 | | | | | 97,50 |
| 4 | (1,1,1,4) | 18,50 | 18,50 | 18,50 | 11,00 | | | | 99,50 |
| | (1,1,2,3) | 14,50 | 14,50 | 13,00 | 10,50 | | | | 86,50 |
| | (1,2,2,2) | 12,50 | 11,00 | 11,00 | 11,00 | | | | 78,50 |
| 5 | (1,1,1,1,3) | 12,50 | 12,50 | 12,50 | 12,50 | 8,50 | | | 75,50 |
| | (1,1,1,2,2) | 10,50 | 10,50 | 10,50 | 9,00 | 9,00 | | | 67,50 |
| | (1,1,1,1,1,2) | 8,50 | 8,50 | 8,50 | 8,50 | 8,50 | 7,00 | | 56,50 |
| 7 | (1,1,1,1,1,1,1) | 6,50 | 6,50 | 6,50 | 6,50 | 6,50 | 6,50 | 6,50 | 45,50 |

Table 2 : Individual and total payoff according to the agreement structure

| Agreement structure | Total payoff | Gini index | Veto | dictator |
|---|---|---|---|---|
| (7) | 172.00 | 0.0000 | 0.2857 | 0.4609 |
| (1,6) | 151.00 | 0.0997 | 0.0446 | 0.1217 |
| (2,5) | 137.00 | 0.1099 | 0.0000 | 0.0957 |
| (4,3) | 130.00 | 0.0451 | 0.0000 | 0.0348 |
| (1,1,5) | 126.00 | 0.1366 | 0.0714 | 0.1043 |
| (1,2,4) | 110.50 | 0.1047 | 0.0179 | 0.0174 |
| (1,3,3) | 106.00 | 0.0325 | 0.0268 | 0.0000 |
| (2,2,3) | 97.50 | 0.0440 | 0.0000 | 0.0087 |
| (1,1,1,4) | 99.50 | 0.1292 | 0.0982 | 0.0696 |
| (1,1,2,3) | 86.50 | 0.0743 | 0.0357 | 0.0522 |
| (1,2,2,2) | 78.50 | 0.0164 | 0.0089 | 0.0261 |
| (1,1,1,1,3) | 75.50 | 0.0908 | 0.0804 | 0.0000 |
| (1,1,1,2,2) | 67.50 | 0.0381 | 0.1071 | 0.0000 |
| (1,1,1,1,1,2) | 56.50 | 0.0379 | 0.1339 | 0.0087 |
| (1,1,1,1,1,1,1) | 45.50 | 0.0000 | 0.0893 | 0.0000 |

Table 4 : frequency of agreement structure according to the Gini coefficient

| Source | SS | df | MS | | |
|---|---|---|---|---|---|
| Model | .028273687 | 3 | .009424562 | Number of obs = | 15 |
| Residual | .049658446 | 11 | .004514404 | F( 3, 11) = | 2.09 |
| Total | .077932134 | 14 | .005566581 | Prob > F = | 0.1600 |
| | | | | R-squared = | 0.3628 |
| | | | | Adj R-squared = | 0.1890 |
| | | | | Root MSE = | .06719 |

| freqveto | Coef. | Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| totalpay | .0003446 | .0005302 | 0.650 | 0.529 | -.0008224 | .0015117 |
| gini | -3.338544 | 1.50256 | -2.222 | 0.048 | -6.645656 | -.0314319 |
| gini2 | 20.45931 | 10.92963 | 1.872 | 0.088 | -3.596641 | 44.51526 |
| _cons | .1216331 | .0665864 | 1.827 | 0.095 | -.0249226 | .2681887 |

Regression veto

| Source | SS | df | MS | | |
|---|---|---|---|---|---|
| Model | .146053323 | 3 | .048684441 | Number of obs = | 15 |
| Residual | .044458753 | 11 | .004041705 | F( 3, 11) = | 12.05 |
| Total | .190512076 | 14 | .013608005 | Prob > F = | 0.0008 |
| | | | | R-squared = | 0.7666 |
| | | | | Adj R-squared = | 0.7030 |
| | | | | Root MSE = | .06357 |

| freqdic | Coef. | Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| totalpay | .0024618 | .0005017 | 4.907 | 0.000 | .0013576 | .003566 |
| gini | -4.181379 | 1.421719 | -2.941 | 0.013 | -7.310562 | -1.052195 |
| gini2 | 24.76188 | 10.34159 | 2.394 | 0.036 | 2.000188 | 47.52358 |
| _cons | -.0676457 | .0630039 | -1.074 | 0.306 | -.2063164 | .0710249 |

Regression dict

Table 5 : analysis of total payoff and payoff disparity

| | All periods | 3 final periods |
|---|---|---|
| Global agreement | 0,2857 | 0,4333 |
| at least 3 singleton | 0,5089 | 0,4000 |
| others | 0,2054 | 0,1667 |

(a) Veto

| | All periods | 3 final periods |
|---|---|---|
| Global agreement | 0,4609 | 0,5556 |
| at least 3 singleton | 0,0783 | 0,0741 |
| others | 0,4609 | 0,3704 |

(b) Dictator

Table 3 : Aggregated frequencies of types of agreement structures

| | Type | Complete | singleton |
|---|---|---|---|
| Veto | Fixed | 63.79 | 17.24 |
| | Random | 50.00 | 33.33 |
| dictator | Fixed | 50.67 | 30.67 |
| | Random | 35.00 | 35.00 |

Table 6 : Comparison of initial proposals of extreme agreements for fixed and random groups

| | Subgame Size | | | | | | | Total | Compatible |
|---|---|---|---|---|---|---|---|---|---|
| Remaining players | 7 | 6 | 5 | 4 | 3 | 2 | 1 | Total | Compatible Proposals |
| 7 | **0.422** | 0.055 | 0.042 | 0.025 | 0.055 | 0.101 | **0.300** | 237 | 171 (0.722) |
| 6 | - | **0.366** | 0.030 | 0.015 | 0.075 | 0.097 | **0.418** | 134 | 105 (0.784) |
| 5 | - | - | **0.375** | 0.042 | 0.021 | 0.083 | **0.479** | 96 | 82 (0.854) |
| 4 | - | - | - | **0.372** | 0.053 | 0.181 | **0.394** | 94 | 72 (0.766) |
| 3 | - | - | - | - | **0.382** | 0.147 | **0.471** | 68 | 58 (0.853) |
| | | | | | | | Total | 629 | 488 (0.776) |

*(a) veto treatment*

| | Subgame Size | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Remaining players | 7 | 6 | 5 | 4 | 3 | 2 | 1 | Total | Compatible proposals |
| 7 | **0.461** | 0.026 | 0.043 | 0.026 | 0.009 | 0.122 | **0.313** | 115 | 89 (0.774) |
| 6 | - | **0.306** | 0.111 | 0.000 | 0.056 | 0.000 | **0.528** | 36 | 30 (0.833) |
| 5 | - | - | **0.424** | 0.061 | 0.030 | 0.152 | **0.333** | 33 | 25 (0.757 |
| 4 | - | - | - | **0.750** | 0.083 | 0.167 | **0.000** | 12 | 9 (0.750) |
| 3 | - | - | - | - | **0.600** | 0.000 | **0.400** | 10 | 10 (1.000) |
| | | | | | | | Total | 206 | 163 (0.791) |

*(b) dictator treatment*

Table 7: Proposals and compatible proposals in sub-games (in % of all proposals)

(N.B. subgames of sizes 1 and 2 are omitted because they are necessarily compatible)

| Group | 7 | 6 | 5 | 4 | 3 |
|---|---|---|---|---|---|
| V1 | 0.9024 | 0.8438 | 0.7917 | 1.0000 | 1.0000 |
| V2 | 0.7778 | 0.8889 | 0.8889 | 0.8333 | 1.0000 |
| V3 | 0.4400 | 0.6364 | 0.6667 | 0.7778 | 0.6667 |
| V4 | 0.8333 | 0.6923 | 0.9167 | 0.7857 | 0.8000 |
| V5 | 0.7500 | 1.0000 | 0.7778 | 0.5385 | 0.8000 |
| V6 | 0.6964 | 0.7381 | 0.8636 | 0.6957 | 0.9286 |
| V7 | 0.6818 | 0.6364 | 0.8333 | 1.0000 | 0.7500 |
| V8 | 0.6757 | 1.0000 | 1.0000 | 0.7692 | 0.8182 |
| **Mean** | **0.7197** | **0.8045** | **0.8423** | **0.8000** | **0.8454** |

*(a) veto treatment*

| Group | 7 | 6 | 5 | 4 | 3 |
|---|---|---|---|---|---|
| D9 | 0.9333 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| D10 | 0.8667 | 1.0000 | 1.0000 | 0.6667 | |
| D11 | 0.8000 | 0.7000 | 0.5714 | | 1.0000 |
| D12 | 0.6000 | 0.7500 | 0.6667 | 1.0000 | 1.0000 |
| D13 | 0.6667 | 0.6667 | 0.7500 | 0.5000 | 1.0000 |
| D14 | 0.6000 | 1.0000 | 0.8333 | 0.5000 | 1.0000 |
| D15 | 1.0000 | 1.0000 | 0.6667 | 1.0000 | 1.0000 |
| **Mean** | **0.7810** | **0.8738** | **0.7840** | **0.7778** | **1.0000** |

*(a) dictator treatment*

Table 8 : Proposals of extreme agreements (in %) for each subgame size and each independent group

| **Group** | **7** | | **6** | | **5** | | **4** | | **3** | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *Global* | *Single* | *Largest* | *Single* | *Largest* | *Single* | *Largest* | *Single* | *Largest* | *Single* |
| V1 | 0.4634 | 0.4390 | 0.3750 | 0.4688 | 0.3333 | 0.4583 | 0.3077 | 0.6923 | 0.3000 | 0.7000 |
| V2 | 0.5556 | 0.2222 | 0.4444 | 0.4444 | 0.4444 | 0.4444 | 0.5000 | 0.3333 | 0.0000 | 1.0000 |
| V3 | 0.2400 | 0.2000 | 0.4545 | 0.1818 | 0.3333 | 0.3333 | 0.2222 | 0.5556 | 0.2222 | 0.4444 |
| V4 | 0.3889 | 0.4444 | 0.1538 | 0.5385 | 0.3333 | 0.5833 | 0.3571 | 0.4286 | 0.5000 | 0.3000 |
| V5 | 0.4500 | 0.3000 | 0.1429 | 0.8571 | 0.2222 | 0.5556 | 0.3846 | 0.1538 | 0.4000 | 0.4000 |
| V6 | 0.3393 | 0.3571 | 0.4048 | 0.3333 | 0.3636 | 0.5000 | 0.3478 | 0.3478 | 0.3571 | 0.5714 |
| V7 | 0.5000 | 0.1818 | 0.4545 | 0.1818 | 0.6667 | 0.1667 | 0.6667 | 0.3333 | 0.5000 | 0.2500 |
| V8 | 0.5135 | 0.1622 | 0.3333 | 0.6667 | 0.4545 | 0.5455 | 0.3077 | 0.4615 | 0.5455 | 0.2727 |

*(a) veto treatment*

| **Group** | **7** | | **6** | | **5** | | **4** | | **3** | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *Global* | *Single* | *Largest* | *Single* | *Largest* | *Single* | *Largest* | *Single* | *Largest* | *Single* |
| D9 | 0.7333 | 0.2000 | 0.0000 | 1.0000 | 0.0000 | 1.0000 | 1.0000 | 0.0000 | 1.0000 | 0.0000 |
| D10 | 0.6000 | 0.2667 | 0.2500 | 0.7500 | 0.2500 | 0.7500 | 0.6667 | 0.0000 | - | - |
| D11 | 0.3000 | 0.5000 | 0.2000 | 0.5000 | 0.5714 | 0.0000 | - | - | 0.5000 | 0.5000 |
| D12 | 0.4000 | 0.2000 | 0.2500 | 0.5000 | 0.5000 | 0.1667 | 1.0000 | 0.0000 | 0.3333 | 0.6667 |
| D13 | 0.2667 | 0.4000 | 0.5000 | 0.1667 | 0.5000 | 0.2500 | 0.5000 | 0.0000 | 0.0000 | 1.0000 |
| D14 | 0.4667 | 0.1333 | 0.0000 | 1.0000 | 0.5000 | 0.3333 | 0.5000 | 0.0000 | 1.0000 | 0.0000 |
| D15 | 0.5333 | 0.4667 | 0.5714 | 0.4286 | 0.3333 | 0.3333 | 1.0000 | 0.0000 | 1.0000 | 0.0000 |

*(b) dictator treatment*

Table 9 : Hit rates per independent group of singletons and largest agreement proposals

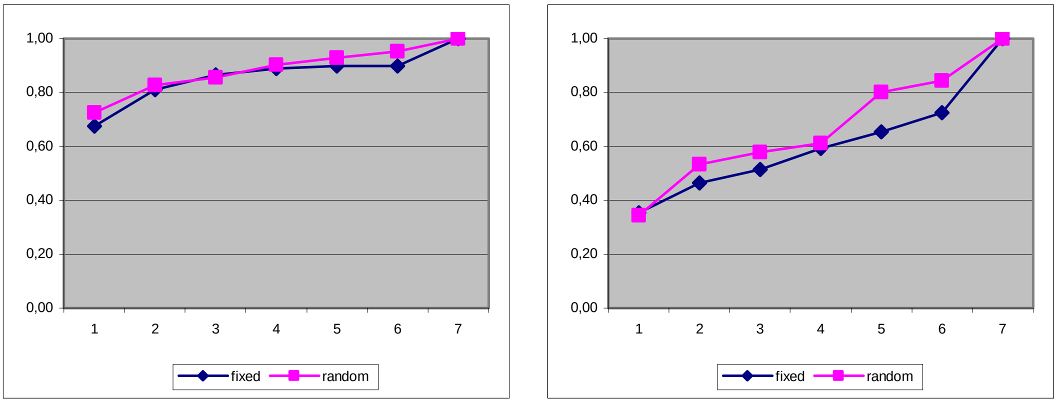
| **Group** | **7** | **6** | **5** | **4** | **3** |
|---|---|---|---|---|---|
| v1 | 0.6167 | 0.5104 | 0.3917 | 0.5000 | 0.3333 |
| v2 | 0.4921 | 0.5556 | 0.4889 | 0.3333 | 0.3333 |
| v3 | 0.1543 | 0.3030 | 0.2667 | 0.2778 | -0.0238 |
| v4 | 0.5476 | 0.3590 | 0.5167 | 0.2857 | 0.1333 |
| v5 | 0.4643 | 0.6667 | 0.3778 | 0.0385 | 0.1333 |
| v6 | 0.4107 | 0.4048 | 0.4636 | 0.1957 | 0.2619 |
| v7 | 0.3961 | 0.3030 | 0.4333 | 0.5000 | 0.0833 |
| v8 | 0.3900 | 0.6667 | 0.6000 | 0.2692 | 0.1515 |
| Mean | **0.4340** | **0.4711** | **0.4423** | **0.3000** | **0.1758** |

*(a) veto treatment*

| **Group** | **7** | **6** | **5** | **4** | **3** |
|---|---|---|---|---|---|
| D9 | 0.6476 | 0.6667 | 0.6000 | 0.5000 | 0.3333 |
| D10 | 0.5810 | 0.6667 | 0.6000 | 0.1667 | |
| D11 | 0.5143 | 0.3667 | 0.1714 | | 0.3333 |
| D12 | 0.3143 | 0.4167 | 0.2667 | 0.5000 | 0.3333 |
| D13 | 0.3810 | 0.3333 | 0.3500 | 0.0000 | 0.3333 |
| D14 | 0.3143 | 0.6667 | 0.4333 | 0.0000 | 0.3333 |
| D15 | 0.7143 | 0.6667 | 0.2667 | 0.5000 | 0.3333 |
| Mean | **0.4952** | **0.5405** | **0.3840** | **0.2381** | **0.2857** |

*(b) dictator treatment*

Table 10 : Measures of predictive success for extreme agreements




(2a) (2b)

Figure 1 : Cumulative frequencies of agreement sizes : veto (2a), dictator (2b)

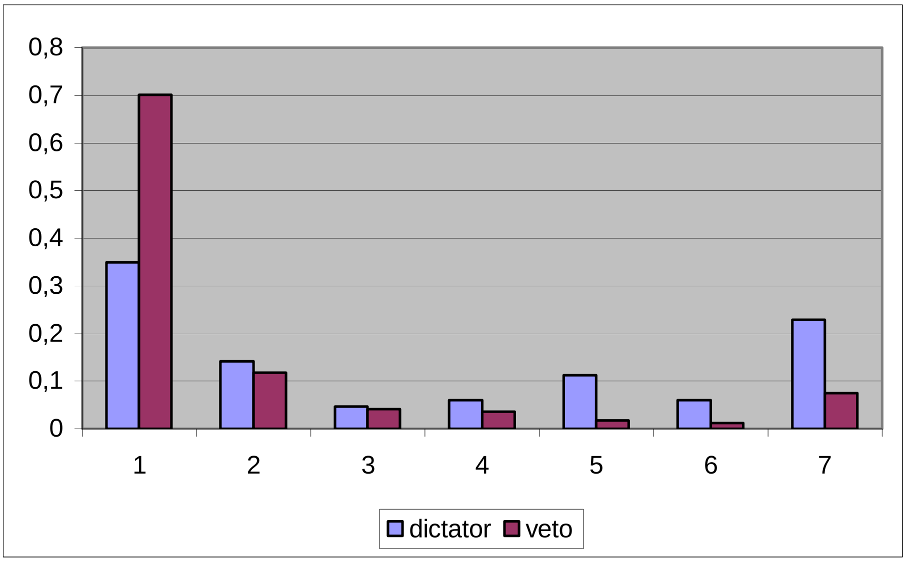



Figure 2 : Size distribution : comparison veto and dictator treatments

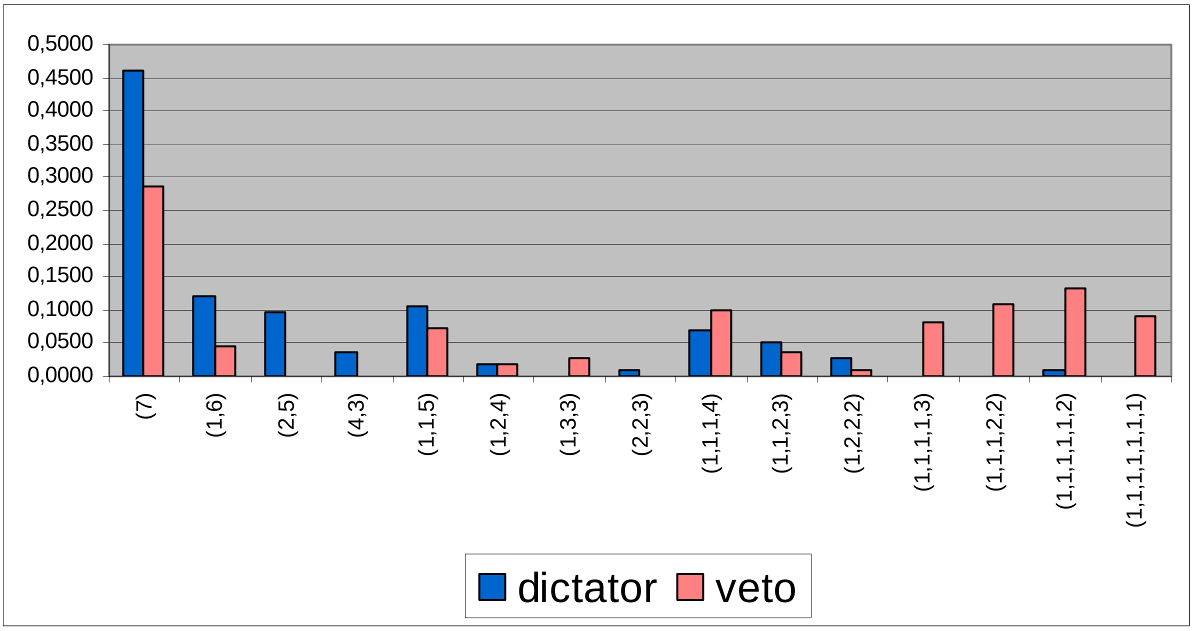



Figure 3 : Frequency of agreements

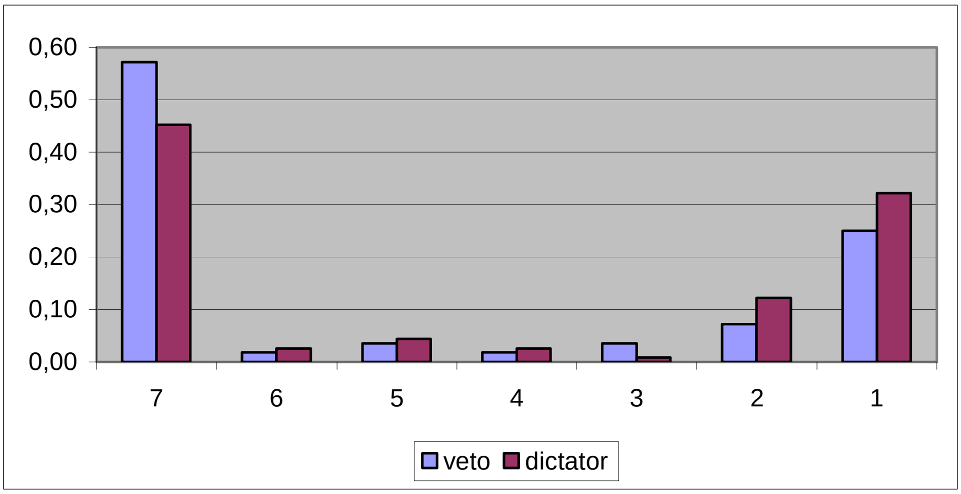



Figure 4 : Frequency distribution of initial proposals (all rounds)

| Structure | $s_1$ | $s_2$ | $s_3$ | $s_4$ | $s_5$ | $s_6$ | $s_7$ | Total payoff |
|---|---|---|---|---|---|---|---|---|
| (7) | 24,50 | | | | | | | 172,00 |
| (1,6) | 36,50 | 19,00 | | | | | | 151,00 |
| (1,1,5) | 26,50 | 26,50 | 14,50 | | | | | 126,00 |
| (1,1,1,4) | 18,50 | 18,50 | 18,50 | 11,00 | | | | 99,50 |
| (1,1,1,1,3) | 12,50 | 12,50 | 12,50 | 12,50 | 8,50 | | | 75,50 |
| (1,1,1,1,1,2) | 8,50 | 8,50 | 8,50 | 8,50 | 8,50 | 7,00 | | 56,50 |
| (1,1,1,1,1,1,1) | 6,50 | 6,50 | 6,50 | 6,50 | 6,50 | 6,50 | 6,50 | 45,50 |

Table 11 : Payoffs in the simplified game