Dichotomous Preferences, Truth-Telling and Collective Action

Marc Vorsatz

Supervisor: Dr. Jordi Massó

Thesis submitted in partial fulfilment of the
requirements for the degree of
Doctor in Economics
at
Universitat Autònoma de Barcelona

April 2005

Contents

To my family and Lena

Acknowledgements

Looking back I realize that a lot of people accompanied and supported me over the last years. Since they all contributed in one or the other way to the results presented here, I would like to take the opportunity to express my indebtedness to all of them.

I would like to express my deepest gratitude to my mentor Jordi Massó for the enthusiasm put in an outstanding supervision. The encouragement I received has been unique, and therefore, his spirit can be found easily in the results of my effort.

I owe special thanks to Santiago Sánchez-Pagés for his collaboration in the fourth chapter, our endless discussions, and his hospitality. Jordi Brandts spent a lot of time explaining me the fundamentals of experimental design. His expertise allowed us to test our experimental hypotheses in a rigorous way. Miguel-Ángel Ballester's excellent comments on the second and third chapter have been crucial to improve my work in a lot of details and to understand it in a better way.

I want to thank all members of the IDEA faculty and the department for their cooperation. Also because of their openness and the friendly environment the last five years have been a very fruitful time for me. Carmen Beviá, Salvador Barberá, Jordi Caballé, Joan Esteban, Laura Mayoral, David Pérez and Hugo Rodríguez have been outstanding teachers in the Master's Program and to them I owe much of what I know today about Economics. Mercè Vicente had always a nice word for me when I stepped into her office and asked desperately for help. I really enjoyed the football talks with Mario Jiménez despite of our different opinions of how to do a good *quiniela*. Finally, I want to mention that this thesis has been financed with the help of the scholarships FI-2000-415 from

Generalitat de Catalunya and BEC 2002-02130 from Ministerio de Ciencia y Tecnología.

Some people apart from my family make me feel home in Germany. I know Dirk and his family already for more than 25 years. Although there is now less time for visits, I am still received as a part of the family. The enthusiasm for sporting events of Eckard and Mario, who taught me the essentials about companionship, is more than remarkable. It has always been a great pleasure to speak with Rita about the philosophy of life and I am still astonished of her convincing judgements.

Being a student at IDEA is a great experience. I met students from a lot of countries with distinct cultural backgrounds and shared with them a very important part of my life. Specially, I want to thank: Gemma, Mayra, Neus and Pau, because without them I would not have come to Barcelona in a first place; Olivia and Vicky for the time we spend together in the Master's Program; Rositza and Atanas for writing the best solutions of problems sets anybody can think of; Héctor, Nico, Marco, Robert and Wendy for being great flatmates (can anybody survive that many changes?); Kaniska for watching football matches at Dow Jones and complaining about Barça's inabilities; everybody with whom I played football and everybody who promised to come and left us waiting afterwards; Edu, Fernanda, Helmuts and Pilar, among others, for all those healthy talks and laughter in our office; Sergio, because he cares about the doctoral program; Joana and Isabel for claiming that my Portuguese is understandable; Ester and Ángel, because we had great dinners together and from them I have learned exemplarily the characteristics of *madrileños* (*madridistas*?); Coralio for having been to clubs I am not entering alone; and, Lena for her love.

Finally, I would like to address some words in honor of my parents. I have been able to arrive to this point because of their education and constant help. They taught me not to fear the obstacles of life.

Bellaterra, Barcelona Marc Vorsatz

April 13 2005

Chapter 1

Introduction

If a group of individuals has to decide upon the selection of some feasible alternatives and individual preferences on the set of alternatives are not aligned, then the institutional problem of how preferences should be aggregated arises. It is the main objective of Social Choice Theory to address this question by studying normative properties of different aggregation rules.

In a lot of socio-economic environments individuals divide alternatives into two indifference classes, the set of good and the set of bad alternatives. According to Roth et al. [50] the set of donors for a potential receiver of a live kidney is dichotomous and depends basically on the blood type of the donor. Similarly, Bogomolnaia et. al [12] claim that individual preferences in time sharing and scheduling problems are dichotomous. As a last example, consider a firm hiring specialized candidates. If the amount of extractable information from the applications is rather low and purchasing external information is expensive, then the recruitment committee members have simple structured preferences. To assume in this setting that preferences are dichotomous is surely not innocuous, yet it constitutes an important benchmark case in the analysis of weak preferences.

The aim of the chapters *Scoring Rules on Dichotomous Preferences* and *Approval Voting on Dichotomous Preferences* is to relate some well known social choice functions and to study their respective properties in order to gain some insight on which rule is best to apply in the kind of situations described above. In Chapter 2, we concentrate on a particular subclass of social choice functions termed scoring rules. The most well known scoring rule, the Borda Count [14], is defined as follows: Given a strict preference relation for some individual, assign zero points to her/his worst alternative, one point to the second worst alternative, and so forth. Then, the alternatives selected for a given preference profile are those with the highest sum of points.

Our main result states that if we generalize the point assignment process to weak preferences, then the Borda Count coincides with Approval Voting [17] on the dichotomous preference domain (Proposition 2.1). Remember that according to Approval Voting every individual can vote for as many alternative as s/he wishes to and for every preference profile all alternatives with the highest number of votes are selected. This result is important, because the two aggregation rules are not easily compared on richer preference domains and there has been an active discussion between Saari and van Newenhizen [52] and [53] and Brams et al. [18] on whether the Borda Count or Approval Voting should be considered the best alternative method for the widely established Plurality Rule (everybody can cast one ballot and all alternatives receiving most votes are implemented). Moreover, Approval Voting coincides on this preference domain with the Condorcet Rule [22] (an alternative is a Condorcet winner if a majority of individuals prefers it to all other alternatives), and therefore, the criticism on the Borda Count not to select an existing Condorcet winner is only valid if preferences consist of at least three indifference classes.

In collective choice problems individuals may try to obtain a better outcome by misrepresenting their preferences. Since strategy-proofness (truth-telling is a dominant strategy in the preference revelation game) eliminates strategic voting completely and insures the selection of the "correct" alternatives, special attention should be given to social choice functions satisfying this normative property. Brams and Fishburn [17] have shown that Approval Voting is strategy-proof on the dichotomous preference domain, a result that opens together with Proposition 2.1 at least two possible lines of investigation. For the remainder Chapter 2, we concentrate on the notion of strategy-proofness and show that any scoring rule different from the Borda Count is manipulable on the dichotomous preference domain (Proposition 2.2) and that there does not exist a domain containing dichotomous preferences for which the Borda Count is strategy-proof whenever the number of individuals participating in the election is different from three (Proposition 2.3). In Chapter 3, on the other hand, we study the normative properties of Approval Voting in more detail and offer two different characterization of this rule. First, we show that on the dichotomous preference domain, a social choice function is anonymous, neutral, strictly monotone and strategy-proof if and only if it is Approval Voting (Proposition 3.1). Afterwards, we characterize Approval Voting on this domain by means of strict symmetry, neutrality and efficiency (Proposition 3.2). These results are related to the axiomatic representations of May [43], Fishburn [35], and Baigent and Xu [3] and support the common opinion that Approval Voting is a very plausible way to aggregate preferences given the domain restriction.

One important aim of the literature on Experimental Economics is to check whether theoretical predictions of economic models withstand tests in the laboratory. In Chapter 4, *An Experimental Study of Truth-Telling in a Sender-Receiver Game* (joint with Santiago Sánchez-Pagés), we show that the rationality assumption underlying the notion of strategy-proofness -voters take any chance to misrepresent their preferences however small the expected gains- can be too strong in certain situations. To do so we analyze in the laboratory a very simple game of "strategic information transmission" [25] and show that individuals have preferences for truth-telling; that is, individuals want to tell the truth about their private information although have incentives to lie. Our first result states that subjects who play the sender-receiver game in the role of the sender tell the truth significantly more often than predicted by the standard model of preference maximization (Hypothesis 4.1). To provide evidence that this "overcommunication phenomenon" of Cai and Wang [20] can be explained because a considerable number of subjects have preferences for truth-telling we test two hypotheses: First, we show that some subjects who have been deceived are willed to pay money in order to punish liars (Hypothesis 4.2). Afterwards, it is established that those subjects who account for most of the punishments tend to tell the truth significantly more than predicted by the standard equilibrium analysis whereas the rest of the subjects play, on the aggregate, equilibrium strategies (Hypothesis

4.3). So subjects do not only behave consistently among roles, rather we can partition our subject pool into two groups, one group with concerns for social preferences and another one following only incentives.

In rent-seeking situations such as lobbying environments individuals spend valuable resources in order to raise the probability to win a fixed prize (see e.g. Tullock [60]). This theoretical finding seems to match empirical observations but bases on the assumption that individuals are not able to cooperate and share the prize without engaging in conflict. The objective of the fifth chapter, *Coalition Formation in a Contest with Three Heterogeneous Players*, is to understand in which situations the efficient outcome, a society-wide agreement, can be achieved and when it is impossible to obtain full cooperation. To address this question we analyze the behavior of three heterogeneous agents in the following two stage model: First, individuals form coalitions according to a bargaining model similar to the partnership game proposed by Gul [39]. Afterwards, the contest game of Esteban and Ray [30] is played in the resulting coalition structure of the first stage. The main result, Proposition 5.2, states that if the relative bargaining power of every individual in the bargaining stage is equal to her/his relative efficiency of lobbying in the contest game, then the efficient grand coalition forms in equilibrium. But, if these two parameters are too distinct, then full cooperation cannot be reached any more, and, as a result, the equilibrium level of effort is strictly positive. Thus, our results contrast the findings of Bloch et al. [10] who show that if individuals are homogenous, then the grand coalition is the unique equilibrium coalition structure.

Chapter 2

Scoring Rules on Dichotomous Preferences

2.1 Introduction

We analyze the aggregation of preferences in form of positional voting methods or scoring rules when individuals have dichotomous preferences on the set of alternatives (there are just two indifference classes, the set of good alternatives and the set of bad alternatives). In particular, we are interested in strategy-proof scoring rules; that is, we look for social choice functions belonging to the class of scoring rules that give individuals incentives to report preferences truthfully. Our main results show that on the dichotomous preference domain, the Borda Count [14] is equal to Approval Voting [17] and the only strategy-proof scoring rule.

In a series of papers, Saari and van Newenhizen [52] and [53] and Brams et al. [18] discuss the advantages and disadvantages of Approval Voting versus scoring rules in general and the Borda Count in particular. The former authors argue that Approval Voting is highly indeterminate for a lot of preference profiles (many different alternatives can be selected for a given preference profile) and suggest the Borda Count as an alternative to the widely established Plurality Rule. But this indeterminacy of Approval Voting is rather a virtue according to the latter authors, because it eliminates voter's incentives not to vote sincerely whereas scoring rules are considered to be very manipulable (see e.g. Dummett [29], Saari [51], and Smith [58]).

One way how to contribute to this discussion is to compare scoring rules with Approval Voting for different preference domains. But this task is not an easy one, because the rules work quite differently. While scoring rules are social choice functions and thus take into account the whole preference structure, Approval Voting is a truncated voting rule that endows individuals with the right to vote for as many alternatives as they wish to and selects all alternatives with the largest support. Therefore, the level of information available about individual preferences is generally lower under Approval Voting. This problem disappears when preferences are restricted to be dichotomous, because if we interpret voting decisions as the set of good alternatives, then individual preferences are fully revealed; that is, Approval Voting becomes a social choice function. Since this is not true any more for larger preference domains, the dichotomous preference domain constitutes an ideal starting point for comparing scoring rules with Approval Voting.

Proposition 2.1 contributes to the former discussion by showing that if the Borda Count is generalized to weak preferences in a straightforward way, then it is an affine transformation of Approval Voting on the domain of dichotomous preferences. Since Brams and Fishburn [17] have shown that Approval Voting is equal to the Condorcet Rule [22] on the dichotomous preference domain (remember that the set of Condorcet winners is non-empty on this domain according to Inada [41]), Proposition 2.1 establishes additionally that the criticism on the Borda Count not to select an existing Condorcet winner is only true if preferences consists of at least three indifference classes.

One common way to eliminate individual incentives to vote strategically is to implement strategy-proof social choice functions. Brams and Fishburn [17] have shown that Approval Voting is strategy-proof on the dichotomous preference domain, and therefore, one may wonder whether other scoring rules share the same property. Since Proposition 2.2 gives a negative answer to this question if there are at least three voters, we can conclude the Borda Count is the best scoring rule in terms of incentives on the dichotomous preference domain.

Finally, we deal with the question of whether we can enlarge the underlying preference domain without losing strategy-proofness for Borda Count. Barbie et al. [6] study strategy-proof domains for the Borda Count under the assumptions that individual preferences are strict and ties are broken in a non-neutral way. Basically, they find that the Borda Count is non-manipulable on all domains which contain one fixed preference relation and all its cyclic permutation. Since these domains are rather small, their result confirms the common opinion that scoring rules are highly manipulable. Proposition 2.3 points into the same direction, because the dichotomous preference domain is the largest domain for which the Borda Count is strategy-proof if more than three individuals participate in the election.

We proceed as follows. In the next section we introduce notation and some basic definitions. Afterwards, we present our results.

2.2 Notation and Definitions

Consider a group of individuals $N = \{1, ..., n\}$ with preferences on the set of alternatives $K = \{1, ..., k\}$. The cardinalities of the two sets are finite and equal to $n \geq 2$ and $k \geq 3$. We assume that $k \geq 3$, because otherwise all scoring rules are going to be equal to the Borda Count as it will become clear from the definitions later on. Elements of K are denoted usually by x, y and z, elements of N by i, j and l.

Let R_i be the weak preference relation of individual i on K. We assume that R_i is reflexive, complete and transitive. The strict and the indifference preference relations associated with R_i are denoted by P_i and I_i, respectively. The set of all weak preference relations on K is denoted by $\mathcal{R}$. A domain $\bar{\mathcal{R}}$ is a subset of $\mathcal{R}$. Given a domain $\bar{\mathcal{R}} \subseteq \mathcal{R}$, a preference profile $R = (R_1, ..., R_n) \in \bar{\mathcal{R}}^N$ is a n-tuple of individual preference relations. The i-variant preference profile (R_i, R_{-i}) is obtained by changing the preference relation of individual i at profile R from R_i to $R_i' \in \bar{\mathcal{R}}$.

The preference relation R_i is dichotomous if it consists of up to two indifference classes, the set of good alternatives and the set of bad alternatives. Given $R_i \in \bar{\mathcal{R}}$, define the set of good alternatives associated with R_i as $G(R_i) = \{x \in K : \; xR_iy \text{ for all } y \in K\}$. Similarly, let $B(R_i) = \{x \in K : \; yR_ix \text{ for all } y \in K\}$ be the set of bad alternatives corresponding to R_i. The cardinalities of these sets are equal to $g(R_i)$ and $b(R_i)$. Hence, $R_i \in \bar{\mathcal{R}}$ is *dichotomous* if and only if $G(R_i) \cup B(R_i) = K$. The domain of all dichotomous preferences is denoted by $\mathcal{D} \subset \mathcal{R}$ and $D_i \in \mathcal{D}$ is a generic dichotomous preference relation for individual i. The notation $G(D_i) = \emptyset$ ($G(D_i) = K$) refers to the situation when i is indifferent between all alternatives and considers no (all) alternative(s) to be acceptable. Finally, given the dichotomous preference profile $D = (D_1, ..., D_n) \in \mathcal{D}^N$, let $N_x(D) = |\{i \in N : x \in G(D_i)\}|$ be the support for x at D.

A social choice function $f : \bar{\mathcal{R}}^N \to 2^K \backslash \{\emptyset\}$ selects for all preference profiles $R \in \bar{\mathcal{R}}^N$ a non-empty set of alternatives $f(R)$. Any social choice function belonging to the class of scoring rules can be represented by a vector $s = (s_0, s_1, \ldots, s_{k-1}) \in \mathbb{R}^k$ satisfying the conditions $s_{j-1} \leq s_j$ for all $j = 1, ..., k-1$ and $s_0 < s_{k-1}$. The range of s is normalized by assuming that $s_0 = 0$ and $s_{k-1} = k-1$. Scoring rules are typically applied to the domain of strict preferences $\mathcal{P}$. In this case, points are assigned to every alternative in such a way that if alternative x is in the j'th position according to P_i, then x receives $p_x^s(P_i) = s_{k-j}$ points from individual i. Given a preference profile $P \in \mathcal{P}^N$ and an alternative $x \in K$, let $p_x^s(P) = \sum_{i=1}^n p_x^s(P_i)$ be the score of alternative x at P when the scoring rule s is applied. Society selects for a given preference profile the set of alternatives with the highest score.

However, if preferences are not strict, then the point assignment process has to be generalized. One possibility is to give to every alternative of the same indifference class the same amount of points, an extension mentioned in [52]. Formally, this is done as follows: Let $C^1(R_i)$ be the set of top alternatives for individual i when her/his preference relation is $R_i \in \bar{\mathcal{R}}$. The cardinality of $C^1(R_i)$ is $c^1(R_i)$. Then every alternative $y \in C^1(R_i)$ receives $p_y^s(R_i) = \frac{1}{c^1(R_i)} \sum_{j=1}^{c^1(R_i)} s_{k-j}$ points from individual i. Let $m \geq 2$ and suppose that points have been given to all alternatives contained in the first $m-1$ indifference classes of R_i. Moreover, denote the cardinality of the set of all alternatives contained in the first $m-1$ indifference classes according to R_i by r^{m-1}. Let $C^m(R_i)$ be the set of alternatives belonging to the m'th indifference class of R_i. The cardinality of $C^m(R_i)$

is $c^m(R_i)$. Then, every alternative $z \in C^m(R_i)$ gets $p_z^s(R_i) = \frac{1}{c^m(R_i)} \sum_{j=1}^{c^m(R_i)} s_{(k-r^{m-1}-j)}$ points from individual i. Given a preference profile $R \in \bar{\mathcal{R}}^N$ and an alternative $x \in K$, let $p_x^s(R) = \sum_{i=1}^n p_x^s(R_i)$ be the score of alternative x at R when the scoring rule s is applied. Now, we are able to define scoring rules for all weak preference domains.

Definition 2.1. The social choice function $f_s : \bar{\mathcal{R}}^N \to 2^K \backslash \{\emptyset\}$ associated to the scoring rule s is such that for all $R \in \bar{\mathcal{R}}^N$, $x \in f(R)$ if and only if $p_x^s(R) \geq p_y^s(R)$ for all $y \in K$.

The most well known scoring rule is the Borda Count. It is defined by the vector $s_j = j$ for all $j = 0, ..., k-1$ and denoted by f_B. With a slight abuse of notation we write $p_x(R_i)$ and $p_x(R)$ whenever the Borda Count is applied. Finally, we provide some intuition for the generalized point assignment process corresponding to the Borda Count: Given $R_i \in \bar{\mathcal{R}}$, compare alternative x with every alternative $y \in K \backslash \{x\}$. If xP_iy, then assign one point to x and zero points to y (give the point to y whenever yP_ix). If xI_iy, then split the point equally. The sum of points alternative x obtains after performing all pair-wise comparisons is equal to $p_x(R_i)$.

Approval Voting is one of the most prominent voting rules both in theory and practice. The main novelty with respect to the Plurality Rule is that it endows individuals with the right to vote for not just one but for as many alternatives as they wish to. That is, the mapping $M_i : \bar{\mathcal{R}} \to 2^K$ determines for all preference relations $R_i \in \bar{\mathcal{R}}$ the set of alternatives $M_i(R_i) \in 2^K$ individual i votes for and the Approval Voting function $v : \left(2^K\right)^N \to 2^K \backslash \{\emptyset\}$ aggregates the individual voting decisions by selecting the alternatives with the highest number of votes. Hence, for all $(M_1(R_1), \ldots, M_n(R_n)) \in \left(2^K\right)^N$, $x \in v\left(M_1(R_1), \ldots, M_n(R_n)\right)$ if and only if $|\{i \in N : x \in M_i(R_i)\}| \geq |\{i \in N : y \in M_i(R_i)\}|$ for all $y \in K$. In [49] there is a discussion of different probabilistic models that make assumptions on how the mappings $(M_i)_{i \in N}$ look like in order to compare Approval Voting in expected terms to social choice functions. But for the case of dichotomous preferences there is a simpler way how to do this. If the mappings $M_i : \mathcal{D} \to 2^K$ are defined such that for all $i \in N$ and all $D_i \in \mathcal{D}$, $M_i(D_i) = G(D_i)$, then the voting decisions reveal preferences completely. Thus, Approval Voting becomes a social choice function.

Definition 2.2. The social choice function $f : \mathcal{D}^N \to 2^K \backslash \{\emptyset\}$ is *Approval Voting* if for all $D \in \mathcal{D}^N$, $x \in f(D)$ if and only if $N_x(D) \geq N_y(D)$ for all $y \in K$.

The social choice function corresponding to Approval Voting is denoted by f_A. Approval Voting is not a social choice function any more if preferences are richer, because, given the voting decision $M_i(R_i)$ for a particular preference relation $R_i \in \bar{\mathcal{R}}$ that consists of at least three indifference classes, it is impossible to recover preferences fully. To see this let the preference relation of individual i be such that xP_iyP_iz. In this case, we cannot infer from $M_i(R_i) = \{x, y\}$ that xP_iy. Similarly, if $M_i(R_i) = \{x\}$, then we cannot deduce that yP_iz.

2.3 Results

The dichotomous preference domain is a natural starting point for a comparison of Approval Voting and scoring rules, because it is the largest domain on which Approval Voting constitutes a well-defined social choice function. Proposition 2.1 states that the Borda Count is equivalent to Approval Voting on this domain, because the score of an alternative under the Borda Count is an affine transformation of the number of individuals who approve it.

Proposition 2.1. *For all $D \in \mathcal{D}^N$, $f_B(D) = f_A(D)$.*

Proof. Suppose that i's preferences are represented by the dichotomous preference relation D_i and let the scoring rule s be such that $s_j = j$ for all $j = 0, ..., k-1$. We deduce from the equation $\sum_{j=1}^{n} j = \frac{n(n+1)}{2}$ that every alternative $x \in G(D_i)$ receives

$$p_x(D_i) = \frac{\sum_{j=1}^{g(D_i)} k-j}{g(D_i)} = \frac{g(D_i)k-\sum_{j=1}^{g(D_i)} j}{g(D_i)} = \frac{2g(D_i)k-g(D_i)(g(D_i)+1)}{2g(D_i)} = \frac{2k-g(D_i)-1}{2}$$

points from individual i. Similarly, this individual gives to all alternatives $y \in B(D_i)$,

$$p_y(D_i) = \frac{\sum_{j=1}^{b(D_i)} k-g(D_i)-j}{b(D_i)} = \frac{2b(D_i)(k-g(D_i))-b(D_i)(b(D_i)+1)}{2b(D_i)} = \frac{k-g(D_i)-1}{2}$$

points. We complete the proof by showing that, given a preference profile $D \in \mathcal{D}^N$ and an alternative $x \in K$, the score $p_x(D)$ is an increasing function of $N_x(D)$. This is done as follows,

$$\begin{aligned} p_x(D) = & \sum_{i \in N: x \in G(D_i)} p_x(D_i) + \sum_{i \in N: x \notin G(D_i)} p_x(D_i) \\ = & \sum_{i \in N: x \in G(D_i)} \tfrac{2k-g(D_i)-1}{2} + \sum_{i \in N: x \notin G(D_i)} \tfrac{k-g(D_i)-1}{2} \\ = & \sum_{i \in N: x \in G(D_i)} \tfrac{k}{2} + \sum_{i \in N} \tfrac{k-g(D_i)-1}{2} \\ = & \tfrac{k}{2} N_x(D) + \tfrac{n}{2}(k-1) - \sum_{i \in N} \tfrac{g(D_i)}{2}. \end{aligned}$$

Hence, for all $D \in \mathcal{D}^N$, $N_x(D) \geq N_y(D)$ for all $y \in K$ if and only if $p_x(D) \geq p_y(D)$ for all $y \in K$. □

One aim of the literature on social choice theory is to study normative properties of aggregation functions. Special attention should be given to strategy-proof social choice functions, because they assure that individuals have incentives to represent their preferences truthfully, or, to say it differently, all room for strategic voting is eliminated. So far, we cannot define strategy-proofness properly, because we do not know how individuals compare non-empty subsets of alternative. One way to deal with this problem is to extend preferences. In particular, we assume that the reflexive, complete and transitive preference relation $\succsim_{R_i}$ defined on $2^K \backslash \{\emptyset\}$ satisfies the subsequent properties proposed by Brams and Fishburn [17]:

1. *Condition* P: $\{x\} \succ_{R_i} \{x, y\} \succ_{R_i} \{y\}$ if and only if $x \in G(R_i)$ and $y \in B(R_i)$.
2. *Condition* R: For all $S, T \subseteq 2^K \backslash \{\emptyset\}$, if $S \subseteq G(R_i)$ or $T \subseteq B(R_i)$ or $[S \backslash T \subseteq G(R_i)$ and $T \backslash S \subseteq B(R_i)]$, then $S \succsim_{R_i} T$.

Now, we can define strategy-proofness in a straightforward way. The social choice function $f : \bar{\mathcal{R}}^N \to 2^K \backslash \{\emptyset\}$ is *manipulable by* i *on* $\bar{\mathcal{R}}^N$ if for some $R \in \bar{\mathcal{R}}^N$ and $R_i' \in \bar{\mathcal{R}}$, $f(R_i', R_{-i}) \succ_{R_i} f(R)$.

Definition 2.3. The social choice function $f : \bar{\mathcal{R}}^N \to 2^K \backslash \{\emptyset\}$ is *strategy-proof on* $\bar{\mathcal{R}}$ if it is not manipulable by any individual on $\bar{\mathcal{R}}^N$.

In [17] it is shown that Approval Voting is strategy-proof on the dichotomous preference domain. According to Proposition 2.2 any scoring rule different from the Borda Count is manipulable on this domain whenever at least three individuals participate in the election.

Proposition 2.2. *Suppose that $n \geq 3$. The social choice function $f_s : \mathcal{D}^N \to 2^K \setminus \{\emptyset\}$ corresponding to the scoring rule s is strategy-proof if and only if it is the Borda Count.*

Proof. The Borda Count is strategy-proof on $\mathcal{D}$ according to Brams and Fishburn [17] and Proposition 2.1. Thus, we only have to show that all scoring rules different from the Borda Count are manipulable on this preference domain. To do so we construct a set of necessary conditions and show that only the Borda Count satisfies them.

Given $x, y \in K$, consider the preference profile $D \in \mathcal{D}^N$ which is as follows: For $i, j \in N$, $G(D_i) = \{x\}$ and $G(D_j) = \{y\}$. For all individuals $l \neq i, j$, D_l is the dichotomous preference relation where $G(D_l) = \{x, y\}$. Then, for any scoring rule s, $f_s(D) = \{x, y\}$. We analyze under which conditions individual i may not manipulate f_s at D via D'_i, where D'_i satisfies the conditions $g(D'_i) > 1$, $x \in G(D'_i)$ and $y \notin G(D'_i)$.

Let $m = g(D'_i) - 1$ be the difference in the cardinality of the set of good alternatives with respect to the preference relations D'_i and D_i. At $(D'_i, D_{-i}) \in \mathcal{D}^N$, the score of alternative x is equal to

$$p_x^s(D'_i, D_{-i}) = \frac{\sum_{j=k-m-1}^{k-1} s_j}{m+1} + \frac{\sum_{j=1}^{k-2} s_j}{k-1} + (n-2)\frac{s_{k-1} + s_{k-2}}{2},$$

because $p_x^s(D'_i) = \frac{1}{m+1}\sum_{j=1}^{m+1} s_{k-j}$, $p_x^s(D_j) = \frac{1}{k-1}\sum_{j=1}^{k-1} s_{k-1-j}$ and $p_x^s(D_l) = \frac{1}{2}(s_{k-1} + s_{k-2})$ for all $l \neq i, j$. At the same preference profile the score of alternative y is equal to

$$p_y^s(D'_i, D_{-i}) = \frac{\sum_{j=1}^{k-m-2} s_j}{k-m-1} + s_{k-1} + (n-2)\frac{s_{k-1} + s_{k-2}}{2},$$

because $p_y^s(D'_i) = \frac{1}{k-m-1}\sum_{j=1}^{k-m-1} s_{k-(m+1)-j}$, $p_y^s(D_j) = s_{k-1}$ and $p_y^s(D_l) = \frac{1}{2}(s_{k-1} + s_{k-2})$ for all $l \neq i, j$. Since the score of alternative x at $(D'_i, D_{-i}) \in \mathcal{D}^N$ is higher as the score of any alternative $z \neq y$ (here we need the assumption $n \geq 3$), individual i cannot manipulate f_s at $D \in \mathcal{D}^N$ via $D_i' \in \mathcal{D}$ whenever $p_x(D_i', D_{-i}) \leq p_y(D_i', D_{-i})$, or for all $m = 1, ..., k-2$,

$$\frac{\sum_{j=k-m-1}^{k-1} s_j}{m+1} + \frac{\sum_{j=1}^{k-2} s_j}{k-1} \leq s_{k-1} + \frac{\sum_{j=1}^{k-m-2} s_j}{k-m-1}.$$

On the other hand, if the former weak inequality is strict for some m, then i can manipulate f_s at $(D_i', D_{-i}) \in \mathcal{D}^N$ via $D_i \in \mathcal{D}$. Hence, the set of equations

$$\frac{\sum_{j=k-m-1}^{k-1} s_j}{m+1} + \frac{\sum_{j=1}^{k-2} s_j}{k-1} = s_{k-1} + \frac{\sum_{j=1}^{k-m-2} s_j}{k-m-1}, \tag{2.1}$$

for all $m = 1, ..., k-2$, defines a set of necessary conditions for strategy-proofness given the generic social choice function f_s. Since the Borda Count is strategy-proof on the dichotomous preference domain, we already know that it solves the linear system of $k-2$ equations (the possible deviations $m = 1, ..., k-2$) and $k-2$ unknowns (the scores s_j for all $j = 1, ..., k-2$). Nonetheless, we present the calculus of the Borda Count before showing that the system (2.1) has a unique solution.

Suppose that $s_j = j$ for all $j = 1, \ldots, k-2$. We have to verify that the equation

$$\frac{\sum_{j=k-m-1}^{k-1} j}{m+1} + \frac{\sum_{j=1}^{k-2} j}{k-1} = k - 1 + \frac{\sum_{j=1}^{k-m-2} j}{k-m-1}$$

holds. Rewrite it as

$$\frac{\sum_{j=1}^{k-1} j - \sum_{j=1}^{k-m-2} j}{m+1} = k - 1 + \frac{\sum_{j=1}^{k-m-2} j}{k-m-1} - \frac{\sum_{j=1}^{k-2} j}{k-1}$$

and apply the equation $\sum_{j=1}^{n} j = \frac{n(n+1)}{2}$ in order to get that the left and the right hand sides of the former equation are equal to $\frac{(k-1)k-(k-m-2)(k-m-1)}{2(m+1)}$ and $\frac{2(k-1)+(k-m-2)-(k-2)}{2}$, respectively. Perform all the necessary multiplications to yield the expression

$$\frac{k^2 - k - (k^2 - 2km - 3k + m^2 + 3m + 2)}{2(m+1)} = \frac{2(k-1) - m}{2}.$$

Simplify this equation to $2km + 2k - m^2 - 3m - 2 = (m+1)(2k - m - 2)$. The result follows from simple algebra.

Finally, we prove that there is no other solution to the system of linear equations. Since s_{k-1} and s_0 are normalized to $k-1$ and 0, respectively, we rewrite equation (2.1) for the generic parameter m as

$$\frac{\sum_{j=k-m-1}^{k-2} s_j}{m+1} + \frac{\sum_{j=1}^{k-2} s_j}{k-1} - \frac{\sum_{j=1}^{k-m-2} s_j}{k-m-1} = \frac{m(k-1)}{m+1}.$$

Next, consider the matrix representation $\mathbf{As} = \mathbf{b}$ of the former system of equations where the rows of the matrix $\mathbf{A}$ correspond to the different values of $m = 1, ..., k-2$. For example, $\mathbf{A} = \left(\frac{1}{k-1}\mathbf{E} + \bar{\mathbf{A}}\right)$, where $\mathbf{E}$ is a $(k-2) \times (k-2)$ matrix with a 1 in every entry and

$$\bar{\mathbf{A}} = \begin{pmatrix} -\frac{1}{k-2} & -\frac{1}{k-2} & \cdots & -\frac{1}{k-2} & \frac{1}{2} \\ -\frac{1}{k-3} & -\frac{1}{k-3} & \cdots & \frac{1}{3} & \frac{1}{3} \\ \vdots & \vdots & & \vdots & \vdots \\ -\frac{1}{2} & \frac{1}{k-2} & \cdots & \frac{1}{k-2} & \frac{1}{k-2} \\ \frac{1}{k-1} & \frac{1}{k-1} & \cdots & \frac{1}{k-1} & \frac{1}{k-1} \end{pmatrix}.$$

Moreover, the vector $\mathbf{b}$ can be represented as $\mathbf{b} = \frac{1}{k-1}\mathbf{b}'$, where

$$\mathbf{b}' = \begin{pmatrix} \frac{1}{2} & \frac{2}{3} & \cdots & \frac{k-3}{k-2} & \frac{k-2}{k-1} \end{pmatrix}.$$

The system of $k-2$ linear equations and $k-2$ unknowns has a unique solution if and only if the matrix of coefficients $\mathbf{A}$ with the generic element $a_{m,r}$ has full rank. Multiply the m'th row of $\mathbf{A}$ by $\frac{1}{a_{m,k-2}} = \left(\frac{1}{k-1} + \frac{1}{m+1}\right)^{-1} = \frac{(k-1)(m+1)}{m+k}$. The resulting matrix $\tilde{\mathbf{A}}$ has the same rank as $\mathbf{A}$ and it is equal to

$$\tilde{\mathbf{A}} = \begin{pmatrix} \tilde{a}_1 & \tilde{a}_1 & \tilde{a}_1 & \ldots & \tilde{a}_1 & \tilde{a}_1 & 1 \\ \tilde{a}_2 & \tilde{a}_2 & \tilde{a}_2 & \ldots & \tilde{a}_2 & 1 & 1 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ \tilde{a}_{k-3} & 1 & 1 & \ldots & 1 & 1 & 1 \\ 1 & 1 & 1 & \ldots & 1 & 1 & 1 \end{pmatrix},$$

where $\tilde{a}_m = \frac{(m+1)(k-1)}{k+m}\left(\frac{1}{k-1} - \frac{1}{k-m-1}\right)$. Observe that for all $m = 1, \ldots, k-3$, $\frac{(m+1)(k-1)}{k+m} > 0$ and $\frac{1}{k-1} - \frac{1}{k-m-1} < 0$. Therefore, $\tilde{a}_m < 0$ for all $m = 1, \ldots, k-3$.

Let $\mathbf{v_r}$ be the vector notation for the r'th column of $\tilde{\mathbf{A}}$. The matrix $\tilde{\mathbf{A}}$ has full rank if and only if there does not exist a vector $\lambda \neq \underline{0}$ such that for all $r = 1, \ldots, k-2$ the scalar product $\lambda \cdot \mathbf{v_r} = 0$. Suppose otherwise; that is, there exists a $\lambda \neq \underline{0}$ such that for all $r = 1, \ldots, k-2$, $\lambda \cdot \mathbf{v_r} = 0$. Consider $\mathbf{v_{k-2}}$ and $\mathbf{v_{k-3}}$. By assumption $\sum_{j=1}^{k-2} \lambda_j = 0$ and $\lambda_1 \tilde{a}_1 + \sum_{j=2}^{k-2} \lambda_j = 0$. Combining the two equations yields $\lambda_1 = \lambda_1 \tilde{a}_1$. Since $\tilde{a}_1 < 0$, it has to be that $\lambda_1 = 0$. Let $m \geq 2$ and suppose that $\lambda_j = 0$ for all $j < m < k-3$. To see that $\lambda_{m+1} = 0$ consider $\mathbf{v_{k-2-m}}$ and $\mathbf{v_{k-3-m}}$. By the induction hypothesis $\sum_{j=m+1}^{k-2} \lambda_j = 0$ and $\lambda_{m+1} \tilde{a}_{m+1} + \sum_{j=m+2}^{k-2} \lambda_j = 0$. Combining the two equations yields $\lambda_{m+1} = \lambda_{m+1} \tilde{a}_{m+1}$. Since $\tilde{a}_{m+1} < 0$, we know that $\lambda_{m+1} = 0$. We conclude that $\lambda_j = 0$ for all $j = 1, ..., k-3$. Finally, since the scalar product $\lambda \cdot \mathbf{v_1} = 0$ by assumption, $\lambda_j = 0$ for all $j = 1, ..., k-3$, and $\tilde{a}_{k-2} = 1 \neq 0$, we conclude that $\lambda_{k-2} = 0$. Hence, $\tilde{\mathbf{A}}$ has full rank. □

The results are mixed for the case when $n = 2$. To see this suppose that $k = 3$. Let the preference profile (D_1, D_2) be such that $G(D_1) = \{x, z\}$ and $G(D_2) = \{y\}$. For this preference profile we find that $f_s(D) = \{y\}$ whenever $s_1 < 1$. Let D_1' be the preference relation corresponding to the set of good alternatives $G(D_1') = \{x\}$. It is easy to see that at (D_1', D_2), $f_s(D_1', D_2) = \{x, y\}$. Since individual 1 with the preference relation D_1 strictly prefers $\{x, y\}$ to $\{y\}$ according to *condition* P, we have found a viable manipulation if $s_1 < 1$. On the other, one can show by means of straightforward calculus that any scoring rule corresponding to the values $s_1 \geq 1$ is strategy-proof on the dichotomous preference domain for some reflexive, complete and transitive preference extension satisfying *condition* P and *condition* R (e.g. the cohesive preferences presented in Chapter 3).

Next, we analyze whether the domain restriction can be weakened, or, to say it differently, whether there are domains containing the set of dichotomous preferences under which the Borda Count is strategy-proof. Following Ching and Serizawa [21], the domain $\bar{\mathcal{R}} \subseteq \mathcal{R}$ is a *maximal domain for a list of properties for the social choice function* $f : \bar{\mathcal{R}}^N \to 2^K \backslash \{\emptyset\}$ if (a) $f : \bar{\mathcal{R}}^N \to 2^K \backslash \{\emptyset\}$ satisfies the list of properties, and (b) for all $\tilde{\mathcal{R}} \supsetneq \bar{\mathcal{R}}$, $f : \tilde{\mathcal{R}}^N \to 2^K \backslash \{\emptyset\}$ does not satisfy the list of properties. In addition to strategy-proofness we consider a richness condition that eliminates all small domains for which the Borda Count is strategy-proof. The condition we apply is stronger than the one of Berga and Serizawa [7] who propose that the domain $\bar{\mathcal{R}} \subseteq \mathcal{R}$ is rich if for all $x \in K$ there exists a preference relation $R_i \in \bar{\mathcal{R}}$ such that xP_iy for all $y \in K \backslash \{x\}$. Here, the domain $\bar{\mathcal{R}} \subseteq \mathcal{R}$

is *rich* if for all $x \in K$ there exists a dichotomous preference relation $D_i \in \bar{\mathcal{R}}$ such that $G(D_i) = \{x\}$. This strengthening is needed, because otherwise we cannot calculate scores and determine the Borda winners. Our final result states that the Borda Count becomes manipulable as we extend the preference domain.

Proposition 2.3. *The dichotomous preference domain is a maximal rich domain for strategy-proofness for the Borda Count. If $n \neq 3$, then it is the unique domain.*

Proof. The proof is organized in two parts. We show at first that the dichotomous preference domain is a maximal rich domain for strategy-proofness for the Borda Count. Afterwards, we proof uniqueness for all cases when $n \neq 3$.

(1) Since for all $x \in K$ the preference relation D_i which is such that $G(D_i) = \{x\}$ belongs to $\mathcal{D}$, it follows that the dichotomous preference domain is rich. Moreover, the Borda Count is strategy-proof on the dichotomous preference domain due to Brams and Fishburn [17] and Proposition 2.1. To see that the Borda Count is manipulable if the underlying preference domain is enlarged, we add the preference relation $T_i \in \mathcal{R}$ with at least three indifference classes to the domain $\mathcal{D}$. Assume without loss of generality that T_i satisfies $C_i^1 = \{1, ..., x^1\}$, $C_i^2 = \{x^1 + 1, ..., x^2\}$, ..., and, $C_i^h = \{x^{h-1} + 1, ..., x^h\}$, where $x^h = k$ and $h \geq 3$. The cardinality of the generic set C_i^j is equal to $|C_i^j| = x^j - x^{j-1} > 0$, where x^0 is normalized to 0. Therefore, we consider now the domains $\tilde{\mathcal{R}} \supseteq \mathcal{D} \cup T_i$ and the objective is to show that the Borda Count is manipulable by i on $\tilde{\mathcal{R}}^N$. Consider the preference profile $D_{-i} \in \tilde{\mathcal{R}}^{N \setminus \{i\}}$ which is such that if $n = 2$, then the preference relation D_j satisfies $G(D_j) = \{x^h\}$. On the other hand, if $n \geq 2$, then let it be such that the preference relation D_j satisfies $G(D_j) = \{x^h\}$ and the preference relation D_l for all $l \neq i, j$ is given by $G(D_l) = \{x^1, x^h\}$. Given the dichotomous preference relation D_i for individual i which corresponds to the set of good alternatives $G(D_i) = \{x^1\}$, we show that individual i can manipulate the Borda Count at $D \in \tilde{\mathcal{R}}^N$ via $T_i \in \tilde{\mathcal{R}}$. At the preference profile (T_i, D_{-i}), $p_{x^h}(D_j) = k - 1$, $p_{x^h}(D_i) = \frac{1}{x^h - x^{h-1}} \sum_{m=1}^{k-x^{h-1}-1} m$, and for all $l \neq i, j$, $p_{x^h}(D_l) = \frac{k-1+k-2}{2}$. Similar at the preference profile (T_i, D_{-i}), $p_{x^1}(D_j) = \frac{1}{k-1} \sum_{m=1}^{k-2} m$, $p_{x^1}(D_i) = \frac{1}{x^1} \sum_{m=k-x^1}^{k-1} m$,

and for all $l \neq i, j,\ p_{x^1}(D_l) = \frac{k-1+k-2}{2}$. Therefore,

$$p_{x^h}(T_i, D_{-i}) = k - 1 + (n-2)\frac{1}{2}(k-1+k-2) + \frac{1}{x^h - x^{h-1}} \sum_{m=1}^{k-x^{h-1}-1} m$$

and

$$p_{x^1}(T_i, D_{-i}) = (n-2)\frac{1}{2}(k-1+k-2) + \frac{1}{x^1} \sum_{m=k-x^1}^{k-1} m + \frac{1}{k-1}\sum_{m=1}^{k-2} m.$$

Since for all $k > j > 0$,

$$\begin{aligned}
\tfrac{1}{x^{j+1}-x^j} \sum_{m=k-x^{j+1}}^{k-x^j-1} m &= \tfrac{1}{x^{j+1}-x^j}\left(\sum_{m=1}^{k-x^j-1} m - \sum_{m=1}^{k-x^{j+1}-1} m\right) \\
&= \tfrac{1}{2}\tfrac{\left(k-x^j-1\right)\left(k-x^j\right)}{x^{j+1}-x^j} - \tfrac{1}{2}\tfrac{\left(k-x^{j+1}-1\right)\left(k-x^{j+1}\right)}{x^{j+1}-x^j} \\
&= \tfrac{k^2-2kx^j+\left(x^j\right)^2-k+x^j}{2(x^{j+1}-x^j)} - \tfrac{k^2-2kx^{j+1}+\left(x^{j+1}\right)^2-k+x^{j+1}}{2(x^{j+1}-x^j)} \\
&= \tfrac{-2kx^j+\left(x^j\right)^2+x^j+2kx^{j+1}-\left(x^{j+1}\right)^2-x^{j+1}}{2(x^{j+1}-x^j)} \\
&= \tfrac{\left(x^{j+1}-x^j\right)(2k-1)-\left(x^{j+1}-x^j\right)\left(x^{j+1}+x^j\right)}{2(x^{j+1}-x^j)} = \tfrac{2k-1-x^{j+1}-x^j}{2}
\end{aligned}$$

it can be concluded that the difference in the score between x^h and x^1 at the preference profile (T_i, D_{-i}), $p_{x^h}(T_i, D_{-i}) - p_{x^1}(T_i, D_{-i})$, is equal to

$$\frac{2\,(k-1)}{2} + \frac{2k-1-x^h-x^{h-1}}{2} - \frac{2k-1-x^1-x^0}{2} - \frac{k-2}{2} = \frac{-x^{h-1}+x^1}{2} < 0.$$

To see this remember that $x^h = k$ and $x^0 = 0$. Therefore, $f_B\,(T_i, D_{-i}) = \{x^1\}$. On the other hand, the score of x^1 and x^h are the same at $D \in \tilde{\mathcal{R}}^N$ which implies that $f_B(D) = \{x^1, x^h\}$. Note that individual i with the preference relation D_i strictly prefers $\{x^1\}$ to $\{x^1, x^h\}$. This is a manipulation, and therefore, the dichotomous preference domain is a maximal rich domain for strategy-proofness for the Borda Count.

(2) Assume that $n \neq 3$. To prove uniqueness, suppose otherwise. Then, there exists a rich domain $\bar{\mathcal{R}} \subseteq \mathcal{R}$ that is not a subset of the dichotomous preference domain and the Borda Count is strategy-proof on the $\bar{\mathcal{R}}$ domain. Since the domain $\bar{\mathcal{R}}$ is rich, given

$x \in K$, the dichotomous preference relation D_i which is such that $G(D_i) = \{x\}$ belongs to $\bar{\mathcal{R}}$. Moreover, since $\bar{\mathcal{R}}$ cannot be a proper subset of $\mathcal{D}$, the preference relation T_i with at least three indifference classes belongs to $\bar{\mathcal{R}}$ as well. Without loss of generality T_i is as described in the first part of the proof. Construct the preference profile $D_{-i} \in \bar{\mathcal{R}}^{N \setminus \{i\}}$ as follows: If n is even, then let there be $\frac{n}{2} - 1$ individuals with the preference relation D_j which is such that $G(D_j) = \{x^1\}$ and $\frac{n}{2}$ individuals with the preference relation D_l which is such that $G(D_l) = \{x^h\}$. If n is odd, then let there be $\frac{n-1}{2} - 1$ individuals with the preference relation D_j which is such that $G(D_j) = \{x^1\}$, $\frac{n-1}{2}$ individuals with the preference relation D_l which is such that $G(D_l) = \{x^h\}$ and one individual $m \in N$ with the preference relation D_m which is such that $G(D_m) = \{x^2\}$. If the preference relation D_i is such that $G(D_i) = \{x^1\}$, then it is easy to see that $f_B(D) = \{x^h, x^1\}$ whenever $n \neq 3$ (to see what happens if $n = 3$ note that $f_B(D) = K$ if $n = k = 3$). Apply the same calculus as in the first part of the proof this Proposition to see that at (T_i, D_{-i}), $f_B(T_i, D_{-i}) = \{x^1\}$ if $n > 3$. Since individual i with the preference relation D_i strictly prefers $\{x^1\}$ to $\{x^1, x^h\}$, the Borda Count is manipulable by i at $D \in \bar{\mathcal{R}}^N$ via $T_i \in \bar{\mathcal{R}}$ whenever $n \neq 3$. Therefore, the domain of dichotomous preferences is the unique maximal rich domain for strategy-proofness for the Borda Count if $n \neq 3$. □

Finally, consider the following example to see why there is another maximal rich domain for strategy-proofness for the Borda Count if the number of individuals is equal to three.

Example 2.1. Suppose that $n = 3$ and $K = \{x, y, z\}$. Let the preference domain $\bar{\mathcal{R}} = \{D_i, D_j, D_l, T_i\}$ be completely prescribed by the sets $G(D_i) = \{x\}, G(D_j) = \{y\}, G(D_l) = \{z\}, G(T_i) = \{x\}$ and $B(T_i) = \{z\}$. Note that the domain $\bar{\mathcal{R}}$ is rich. If the preference profile $R \in \bar{\mathcal{R}}$ is such that two individuals have the same preference relation D_m, $m = i, j, l$, or one individual has the preference relation D_i and a second individual has the preference relation T_i, then the Borda Count selects the top alternative according to D_m or alternative x, respectively. We can see that at these preference profiles the top alternative of two individuals is chosen. Since the third individual cannot change this by misrepresenting her/his preferences, there are only two possible manipulations: Individual i either manipulates the Borda Count at $(D_i, D_j, D_l) \in \bar{\mathcal{R}}^N$ via $T_i \in \bar{\mathcal{R}}$ or s/he manipulates the Borda Count at $(T_i, D_j, D_l) \in \bar{\mathcal{R}}^N$ via $D_i \in \bar{\mathcal{R}}$. Observe that $f_B(D_i, D_j, D_l) = K$ and $f_B(T_i, D_j, D_l) = \{y\}$. If individual i with the preference relation D_i, or T_i respectively, is indifferent between $\{y\}$ and $\{x, y, z\}$ (this does not contradict neither *condition* P nor *condition* R), then the Borda Count is strategy-proof on the $\{D_i, D_j, D_l, T_i\}$ domain. ■

Chapter 3

Approval Voting on Dichotomous Preferences

3.1 Introduction

Our main objective is to study set-valued social choice functions axiomatically when individuals have dichotomous preferences. Unlike standard approaches to such issues, it is not assumed that all individuals necessarily vote, nor that all alternatives are necessarily available. The main results offer two characterizations of Approval Voting [17], one of the most prominent procedures in both theory and practice.

More concretely, we are interested in the following kind of problems: Consider a job offer for specialized candidates. Often, firms decide in a multi-stage procedure whom to contract (e.g. firms invite a number of candidates for an assessment center or a personal interview before taking the final decision), because the amount of extractable information from the applications may be rather low and purchasing external information can be very expensive. In these circumstances, preferences of the recruiting committee members are likely to have a simple structure at the beginning of the decision process. In an extreme case, every member of the recruiting committee classifies candidates either as "acceptable" or as "non-acceptable"; that is, individuals have dichotomous preferences on the set of candidates. The main purpose is to study how the decision makers should aggregate their opinions in this kind of situation and determine the set of pre-selected candidates.

Proposition 3.1 characterizes Approval Voting by means of anonymity, neutrality, strategy-proofness and strict monotonicity. Further axiomatic representations of Approval Voting are due to Fishburn [36] and Sertel [56], but the result which is most closely related to Proposition 3.1 can be found in [35]. There, Fishburn shows that if individuals have dichotomous preferences, then a family of social choice correspondences (the set of alternatives is fixed whereas the set of voters is allowed to vary) is anonymous, neutral, strategy-proof and consistent if and only if it is Approval Voting.[1]

Proposition 3.1 differs from this result mainly because we use strict monotonicity instead of consistency, but the characterizations are nevertheless fundamentally distinct for two reasons. First, consider the following version of May's Theorem [43]: If the number of alternatives is equal to two, then a social choice function is anonymous, neutral and strictly monotone if and only if it is the Majority Rule (Condorcet Rule). Since Brams and Fishburn [17] have shown that Approval Voting is equal to the Condorcet Rule whenever individuals have dichotomous preferences (the set of Condorcet Winners is non-empty on the dichotomous preference domain according to Inada [41]), the main interpretation of Proposition 3.1 is that May's Theorem can be extended to any arbitrary number of alternatives if strategy-proofness is added to the original set of properties. Second, Moulin [47] has pointed out that neutrality, anonymity, strict monotonicity and Independence of Irrelevant Alternatives (IIA) characterize the social welfare function corresponding Approval Voting. Thus, Proposition 3.1 indicates some equivalence between strategy-proofness of a social choice function and IIA of the corresponding social welfare function. Such an equivalence has been formally established for strict preference domains by Blair and Muller [8] but is so far unknown for the dichotomous preference domain.

Afterwards, we show in Proposition 3.2 that Approval Voting is the only strictly symmetric, neutral and efficient social choice function.[2] This result is related to the following

[1] Consistency has the following meaning: If some alternatives are selected for two disjoint electorates, then exactly those alternatives have to be chosen whenever all individuals participate in the election.

[2] Strict symmetry means that the effect on the image of an alternative to be good is independent of who considers this alternative to be acceptable and which other alternatives are acceptable for the vary

characterization of Baigent and Xu [3]: A choice aggregation procedure is neutral, strictly monotone and satisfies Independence of Symmetric Substitution (ISS) if and only if it is Approval Voting. Choice aggregation procedures and social choice functions are generally not comparable, because the domain of the former is the set of all subsets of alternatives (the alternatives an individual votes for) and not preferences. But if preferences are restricted to be dichotomous and we interpret the observed ballots as the set of acceptable alternatives, then voting decisions reveal preferences. In this case, strict symmetry implies ISS, but it turns out that this strengthening is necessary in order to apply efficiency instead of strict monotonicity.

The remainder of the paper is organized as follows. In the next Section, we introduce notation and definitions. The characterizations are presented in the Sections 3 and 4. Afterwards, we conclude. Additional examples can be found in the Appendix.

3.2 Basic Notation and Definitions

Consider a group of individuals N with preferences on the set of alternatives K whose objective is to aggregate their preferences by choosing a non-empty subset of alternatives. Since individuals may abstain from voting, the actual electorate $\bar{N}$ is assumed to be a subset of N. Moreover, it may happen that not all alternatives are feasible, and therefore, we restrict the set of implementable alternatives to be equal to $\bar{K} \subseteq K$. The aggregation problem is interesting only if $|\bar{K}| \equiv \bar{k} \geq 2$ and $|\bar{N}| \equiv \bar{n} \geq 2$.

Let R_i be the weak preference relation of individual i on K. We assume that R_i is reflexive, complete and transitive. The strict and the indifference preference relations associated with R_i are denoted by P_i and I_i, respectively. The set of all weak preferences on K is denoted by $\mathcal{R}$. The preference relation R_i is dichotomous if it consists of up to two indifference classes, the set of good alternatives and the set of bad alternatives. Given $R_i \in \mathcal{R}$, define the set of good alternatives associated with R_i as $G(R_i) = \{x \in K : xR_iy \text{ for all } y \in K\}$. Similarly, let $B(R_i) = \{x \in K : yR_ix \text{ for all } y \in K\}$ be the

same individual.

set of bad alternatives corresponding to R_i. The cardinalities of the two sets are equal to $g(R_i)$ and $b(R_i)$. Then, $R_i \in \mathcal{R}$ is *dichotomous* if and only if $G(R_i) \cup B(R_i) = K$.

The domain of all dichotomous preferences is denoted by $\mathcal{D} \subset \mathcal{R}$ and let $D_i \in \mathcal{D}$ be a particular dichotomous preference relation for individual i. We reserve the notation $G(D_i) = \emptyset$ ($G(D_i) = K$) for the situation when individual i is indifferent between all alternatives and considers no (all) alternative(s) to be acceptable. Given the electorate $\bar{N}$, a preference profile $D_{\bar{N}} = (D_i)_{i \in \bar{N}} \in \mathcal{D}^{\bar{N}}$ is a $\bar{n}$-tuple of dichotomous preference relations. The i-variant preference profile $(D'_i, D_{\bar{N}\setminus\{i\}})$ is obtained by changing the preference relation of individual i in the profile $D_{\bar{N}}$ from D_i to $D'_i \in \mathcal{D}$. Given the preference profile $D_{\bar{N}} \in \mathcal{D}^{\bar{N}}$, let $N(D_{\bar{N}}; x, y)$ be the individuals who weakly prefer x to y at $D_{\bar{N}}$; that is, $N(D_{\bar{N}}; x, y) = \{i \in \bar{N} : x \in G(D_i) \text{ or } x, y \in B(D_i)\}$. Finally, given the preference profile $D_{\bar{N}} \in \mathcal{D}^{\bar{N}}$, $N_x(D_{\bar{N}}) = |\{i \in \bar{N} : x \in G(D_i)\}|$ denotes the support of alternative x at $D_{\bar{N}}$.

A family of social choice functions $\left\{f^{\bar{K},\bar{N}} : \mathcal{D}^{\bar{N}} \to 2^{\bar{K}}\setminus\{\emptyset\}\right\}_{\bar{K},\bar{N}}$ selects for all sets of feasible alternatives $\bar{K}$, all electorates $\bar{N}$, and all preference profiles $D_{\bar{N}} \in \mathcal{D}^{\bar{N}}$ (note that individuals have preferences on K and not on $\bar{K}$) a non-empty set of feasible alternatives $f^{\bar{K},\bar{N}}(D_{\bar{N}})$. With a slight abuse of notation we write $f^{\bar{K}}(D_{\bar{N}})$ instead of $f^{\bar{K},\bar{N}}(D_{\bar{N}})$. Moreover, we suppress indexes throughout whenever no restriction is made on the set of feasible alternatives or the set of individuals.

Two consistency conditions keep track on how the selected set of alternatives varies as the set of feasible alternatives $\bar{K}$ or the electorate $\bar{N}$ changes. The family of social choice functions $\left\{f^{\bar{K},\bar{N}} : \mathcal{D}^{\bar{N}} \to 2^{\bar{K}}\setminus\{\emptyset\}\right\}_{\bar{K},\bar{N}}$ is *consistent in alternatives* if for all sets of feasible alternatives $S \subset T \subseteq K$, all electorates $\bar{N}$, and all preference profiles $D_{\bar{N}} \in \mathcal{D}^{\bar{N}}$, $f^S(D_{\bar{N}}) = f^T(D_{\bar{N}}) \cap S$ whenever $f^T(D_{\bar{N}}) \cap S \neq \emptyset$. Consistency in alternatives has been studied first by Arrow [2] and can be interpreted in the following way: The decision makers suppose a priori that every alternative is feasible and determine which alternatives to pre-select. If it turns out afterwards that fewer alternatives are implementable, then the set of pre-selected alternatives is restricted accordingly.

Given the electorates $A \subset C \subseteq N$ and the preference profile $D_C \in \mathcal{D}^C$, let $D_C|_A \in \mathcal{D}^A$ be the profile obtained by restricting $D_C \in \mathcal{D}^C$ to A. The family of social choice functions

$\left\{f^{\bar{K},\bar{N}} : \mathcal{D}^{\bar{N}} \to 2^{\bar{K}}\backslash\{\emptyset\}\right\}_{\bar{K},\bar{N}}$ is *consistent in individuals* if for all pairs of alternatives (x, y), all electorates $A \subset C \subseteq N$ and all preference profiles $D_A \in \mathcal{D}^A$ and $D_C \in \mathcal{D}^C$ which are such that $D_A = D_C|_A$ and xI_iy for all $i \in C\backslash A$, the condition $f^{\{x,y\}}(D_A) = f^{\{x,y\}}(D_C)$ is satisfied. Hence, consistency in individuals requires that if there are just two feasible alternatives, then individuals who are indifferent between those alternatives cannot alter the result.

A *social choice rule* is a family of social choice functions that is consistent in alternatives and individuals. One particular social choice rule is Approval Voting. According to it all feasible alternatives with the highest support from the electorate are selected.

Definition 3.1. The social choice rule $\left\{f^{\bar{K},\bar{N}} : \mathcal{D}^{\bar{N}} \to 2^{\bar{K}}\backslash\{\emptyset\}\right\}_{\bar{K},\bar{N}}$ is *Approval Voting* if for all sets of feasible alternatives $\bar{K}$, all electorates $\bar{N}$, and all dichotomous preference profiles $D_{\bar{N}} \in \mathcal{D}^{\bar{N}}$, $x \in f^{\bar{K}}(D_{\bar{N}})$ if and only if $N_x(D_{\bar{N}}) \geq N_y(D_{\bar{N}})$ for all $y \in \bar{K}$.

We denote the generic social choice function $f^{\bar{K},\bar{N}}$ associated with Approval Voting by $f_A^{\bar{K},\bar{N}}$. At this point we are ready to introduce the four axioms used in the first characterization of Approval Voting. The first property, strategy-proofness, states that truth-telling is a dominant strategy in the preference revelation game. But since our primitives are social choice correspondences, we have to know how individuals compare non-empty subsets of alternatives in order to define strategy-proofness properly. In particular, we assume that the reflexive, complete and transitive preference relation $\succsim_{D_i}$ on $2^K\backslash\{\emptyset\}$ derived from the dichotomous preference relation D_i satisfies the subsequent properties proposed by Brams and Fishburn [17]:

1. *Condition* P: $\{x\} \succ_{D_i} \{x, y\} \succ_{D_i} \{y\}$ if and only if $x \in G(D_i)$ and $y \in B(D_i)$.

2. *Condition* R: For all $S, T \subseteq 2^K\backslash\{\emptyset\}$, if $S \subseteq G(D_i)$ or $T \subseteq B(D_i)$ or $[S\backslash T \subseteq G(D_i)$ and $T\backslash S \subseteq B(D_i)]$, then $S \succsim_{D_i} T$.

Now, we can define strategy-proofness in a straightforward way. Given the set of feasible alternatives $\bar{K}$ and the electorate $\bar{N}$, the social choice function $f^{\bar{K},\bar{N}} : \mathcal{D}^{\bar{N}} \to 2^{\bar{K}}\backslash\{\emptyset\}$ is *manipulable by* i if for some $D_{\bar{N}} \in \mathcal{D}^{\bar{N}}$ and $D'_i \in \mathcal{D}$, $f^{\bar{K}}(D'_i, D_{\bar{N}\backslash\{i\}}) \succ_{D_i} f^{\bar{K}}(D_{\bar{N}})$.

Definition 3.2. The social choice rule $\left\{f^{\bar{K},\bar{N}} : \mathcal{D}^{\bar{N}} \to 2^{\bar{K}} \backslash \{\emptyset\}\right\}_{\bar{K},\bar{N}}$ is said to be *strategy-proof* if for all sets of feasible alternatives $\bar{K}$ and all electorates $\bar{N}$, $f^{\bar{K},\bar{N}}$ is not manipulable by any individual.

Anonymity (Neutrality) formalizes the democratic idea that there is no a priori bias in favor of some individual (alternative). Given the preference profile $D_{\bar{N}} \in \mathcal{D}^{\bar{N}}$ and the permutation σ of $\bar{N}$, let $D_{\sigma(\bar{N})} \in \mathcal{D}^{\bar{N}}$ be the preference profile obtained by permuting individuals according to σ; that is, $D_{\sigma(\bar{N})} = \left(D_{\sigma(i)}\right)_{i \in \bar{N}}$.

Definition 3.3. The social choice rule $\left\{f^{\bar{K},\bar{N}} : \mathcal{D}^{\bar{N}} \to 2^{\bar{K}} \backslash \{\emptyset\}\right\}_{\bar{K},\bar{N}}$ is said to be *anonymous* if for all sets of feasible alternatives $\bar{K}$, all electorates $\bar{N}$, all preference profiles $D_{\bar{N}} \in \mathcal{D}^{\bar{N}}$, and all permutations σ of $\bar{N}$, $f^{\bar{K}}(D_{\sigma(\bar{N})}) = f^{\bar{K}}(D_{\bar{N}})$.

Given the preference profile $D_{\bar{N}} \in \mathcal{D}^{\bar{N}}$ and the permutation μ of K, $\mu(D_{\bar{N}}) \in \mathcal{D}^{\bar{N}}$ is the preference profile obtained by permuting alternatives according to μ; that is, for all $i \in \bar{N}$ and $x \in K$, $x \in \mu(G(D_i))$ if and only if $\mu^{-1}(x) \in G(D_i)$.

Definition 3.4. The social choice rule $\left\{f^{\bar{K},\bar{N}} : \mathcal{D}^{\bar{N}} \to 2^{\bar{K}} \backslash \{\emptyset\}\right\}_{\bar{K},\bar{N}}$ is said to be *neutral* if for all sets of feasible alternatives $\bar{K}$, all electorates $\bar{N}$, all permutations μ of K, and all preference profiles $D_{\bar{N}} \in \mathcal{D}^{\bar{N}}$, $f^{\bar{K}}(\mu(D_{\bar{N}})) = \mu\left(f^{\bar{K}}(D_{\bar{N}})\right)$.

In the former definition, the set $\mu\left(f^{\bar{K}}(D_{\bar{N}})\right)$ is obtained by applying μ to $f^{\bar{K}}(D_{\bar{N}})$. The last property to be introduced is strict monotonicity. To explain it suppose that x and y are the only feasible alternatives and that both alternatives are selected for a particular preference profile. The tie occurring in this situation should then be broken in favor of x whenever this alternative receives additional support everything else unchanged.

Definition 3.5. The social choice rule $\left\{f^{\bar{K},\bar{N}} : \mathcal{D}^{\bar{N}} \to 2^{\bar{K}} \backslash \{\emptyset\}\right\}_{\bar{K},\bar{N}}$ is said to be *strictly monotone* if for all pairs of alternatives (x, y), all electorates $\bar{N}$, and all preference profiles $D_{\bar{N}}, D'_{\bar{N}} \in \mathcal{D}^{\bar{N}}$ which are such that $G(D'_i) = G(D_i) \cup \{x\}$ and $x \notin G(D_i)$ for some $i \in \bar{N}$ and $D_{\bar{N}\backslash\{i\}} = D'_{\bar{N}\backslash\{i\}}$, the condition $x \in f^{\{x,y\}}(D_{\bar{N}})$ implies $f^{\{x,y\}}(D'_{\bar{N}}) = \{x\}$.

3.3 A Characterization with Strategy-Proofness

Our main result characterizes Approval Voting by means of strategy-proofness, anonymity, neutrality and strict monotonicity. The proof we provide is organized along an important lemma showing that if a social choice rule is neutral and strategy-proof, then it depends on the individuals who prefer x to y and y to x whenever there are no other feasible alternatives. Formally, the social choice rule $\left\{f^{\bar{K},\bar{N}} : \mathcal{D}^{\bar{N}} \to 2^{\bar{K}}\backslash\{\emptyset\}\right\}_{\bar{K},\bar{N}}$ satisfies *Independence of Irrelevant Alternatives* (IIA) if for all pairs of alternatives (x,y), all electorates $\bar{N}$, and all preference profiles $D_{\bar{N}}, D'_{\bar{N}} \in \mathcal{D}^{\bar{N}}$ which are such that $N(D_{\bar{N}}; x, y) = N(D'_{\bar{N}}; x, y)$ and $N(D_{\bar{N}}; y, x) = N(D'_{\bar{N}}; y, x)$, the condition $f^{\{x,y\}}(D_{\bar{N}}) = f^{\{x,y\}}(D'_{\bar{N}})$ holds.

Lemma 3.1. *If the social choice rule* $\left\{f^{\bar{K},\bar{N}} : \mathcal{D}^{\bar{N}} \to 2^{\bar{K}}\backslash\{\emptyset\}\right\}_{\bar{K},\bar{N}}$ *is neutral and strategy-proof, then it satisfies IIA.*

Proof. Suppose otherwise. Then, there is an electorate $\bar{N}$ and two preference profiles $D_{\bar{N}}, D'_{\bar{N}} \in \mathcal{D}^{\bar{N}}$ which are such that for some pair of alternatives (x,y), $N(D_{\bar{N}}; x, y) = N\left(D'_{\bar{N}}; x, y\right)$ and $N(D_{\bar{N}}; y, x) = N\left(D'_{\bar{N}}; y, x\right)$, whereas the social choice function $f^{\{x,y\},\bar{N}}$ satisfies $f^{\{x,y\}}(D_{\bar{N}}) = \{x\}$ and $f^{\{x,y\}}\left(D'_{\bar{N}}\right) \in \{\{y\},\{x,y\}\}$. Let $i \in C \subseteq \bar{N}$ if and only if xP_iy or yP_ix. If $C = \emptyset$, then xI_iy for all $i \in \bar{N}$. In this case, it has to be that $f^{\{x,y\}}(D_{\bar{N}}) = \{x,y\}$, because the function $f^{\{x,y\},\bar{N}}$ is neutral by assumption and the empty set cannot be selected. This is a contradiction to $f^{\{x,y\}}(D_{\bar{N}}) = \{x\}$, and therefore, $C \neq \emptyset$. Now, apply consistency in individuals to obtain that $f^{\{x,y\}}\left(D_{\bar{N}}|_C\right) = \{x\}$. For simplicity let the preference profile $D_C \in \mathcal{D}^C$ be such that $D_C = D_{\bar{N}}|_C$.

At first we prove that for all $j \in C$, $f^{\{x,y\}}\left(D'_j, D_{C\backslash\{j\}}\right) = \{x\}$. Suppose otherwise; that is, $f^{\{x,y\}}\left(D'_j, D_{C\backslash\{j\}}\right) \in \{\{y\},\{x,y\}\}$. If xP_iy, then j can manipulate $f^{\{x,y\},C}$ at $\left(D'_j, D_{C\backslash\{j\}}\right) \in \mathcal{D}^C$ via $D_j \in \mathcal{D}$. On the other hand, if yP_ix, then j can manipulate $f^{\{x,y\},C}$ at $D_C \in \mathcal{D}^C$ via $D'_j \in \mathcal{D}$. This is a contradiction, and therefore, we can conclude that $f^{\{x,y\}}\left(D'_j, D_{C\backslash\{j\}}\right) = \{x\}$.

Let $M \subset C$ be such that $j \in M$, $2 \leq |M| < |C|$, and suppose that for all $\hat{M} \subseteq M$ satisfying $j \in \hat{M}$, $f^{\{x,y\}}\left(D'_{\hat{M}}, D_{C\backslash\hat{M}}\right) = \{x\}$. We have to proof that for all $\tilde{M} = M \cup \{i\}$,

$i \in C\backslash M$, $f^{\{x,y\}}\left(D'_{\tilde{M}}, D_{C\backslash\tilde{M}}\right) = \{x\}$. Suppose otherwise, that is $f^{\{x,y\}}\left(D'_{\tilde{M}}, D_{C\backslash\tilde{M}}\right) \in \{\{y\},\{x,y\}\}$. If xP_iy, then i can manipulate $f^{\{x,y\},C}$ at $\left(D'_{\tilde{M}}, D_{C\backslash\tilde{M}}\right) \in \mathcal{D}^C$ via $D_i \in \mathcal{D}$. On the other hand, if yP_ix, then i can manipulate $f^{\{x,y\},C}$ at $\left(D_M, D_{C\backslash M}\right) \in \mathcal{D}^C$ via $D'_i \in \mathcal{D}^C$. Hence, it has to be that $f^{\{x,y\}}\left(D'_{\tilde{M}}, D_{C\backslash\tilde{M}}\right) = \{x\}$. In particular, if $M = C\backslash\{i\}$, then $f^{\{x,y\}}\left(D'_{\tilde{M}}, D_{C\backslash\tilde{M}}\right) = f^{\{x,y\}}\left(D'_C\right) = \{x\}$. Finally, $f^{\{x,y\}}\left(D'_C, D'_{\bar{N}\backslash C}\right) = f^{\{x,y\}}\left(D'_{\bar{N}}\right) = \{x\}$, because the family $\left\{f^{\{x,y\},\bar{N}}\right\}_{\bar{N}}$ is consistent in individuals and xI'_iy for all $i \in \bar{N}\backslash C$ (to see this remember that the preference profiles $D_{\bar{N}}, D'_{\bar{N}} \in \mathcal{D}^{\bar{N}}$ satisfy the conditions $N(D_{\bar{N}}; x, y) = N\left(D'_{\bar{N}}; x, y\right)$ and $N(D_{\bar{N}}; y, x) = N\left(D'_{\bar{N}}; y, x\right)$; thus, $xI_iy \Leftrightarrow xI'_iy$). This contradicts the assumption $y \in f^{\{x,y\}}(D'_{\bar{N}})$. □

We illustrate why it is not possible to dispense of consistency in individuals in Lemma 3.1 before stating our main result. Example 3.1 presents a family of social choice functions that is neutral, strategy-proof and consistent in alternatives but fails to satisfy IIA.

Example 3.1. The family of social choice functions $\left\{f^{\bar{K},\bar{N}} : \mathcal{D}^{\bar{N}} \to 2^{\bar{K}}\backslash\{\emptyset\}\right\}_{\bar{K},\bar{N}}$ is as follows: For all sets of feasible alternatives $\bar{K}$, all electorates $\bar{N}$, and all preference profiles $D_{\bar{N}} \in \mathcal{D}^{\bar{N}}$, if there is an $i \in \bar{N}$ whose preference relation D_i is such that $G(D_i) = K$, then $f^{\bar{K}}(D_{\bar{N}}) = \bar{K}$. Otherwise, $f^{\bar{K}}(D_{\bar{N}}) = f^{\bar{K}}_A(D_{\bar{N}})$. The family $\left\{f^{\bar{K},\bar{N}}\right\}_{\bar{K},\bar{N}}$ is neutral, strategy-proof (there are no incentives either to vote for a bad alternative or not to vote for a good alternative) and consistent in alternatives. To see that this family does not satisfy IIA consider the case where $K = \{x, y, z\}$, $N = \{1, 2\}$ and the preference profiles $D, D' \in \mathcal{D}^N$ are such that $G(D_1) = \{x, y\}$, $G(D'_1) = K$ and $G(D_2) = G(D'_2) = \{x\}$. Then, $f^{\{x,y\}}(D) = \{x\}$ and $f^{\{x,y\}}(D') = \{x, y\}$. This contradicts IIA. ■

Proposition 3.1. *The social choice rule* $\left\{f^{\bar{K},\bar{N}} : \mathcal{D}^{\bar{N}} \to 2^{\bar{K}}\backslash\{\emptyset\}\right\}_{\bar{K},\bar{N}}$ *is strategy-proof, neutral, anonymous and strictly monotone if and only if it is Approval Voting.*

Proof. Observe that the social choice rule corresponding to Approval Voting is neutral, anonymous, strictly monotone and strategy-proof. To prove the other inclusion suppose that $\left\{f^{\bar{K},\bar{N}} : \mathcal{D}^{\bar{N}} \to 2^{\bar{K}}\backslash\{\emptyset\}\right\}_{\bar{K},\bar{N}}$ satisfies the four properties. At first we show that the family of social choice functions $\left\{f^{\{x,y\},\bar{N}} : \mathcal{D}^{\bar{N}} \to 2^{\{x,y\}}\backslash\{\emptyset\}\right\}_{\bar{N}}$ orders x and y according to Approval Voting; that is, for all electorates $\bar{N}$ and all preference profiles $D_{\bar{N}} \in \mathcal{D}^{\bar{N}}$, (a) if $N_x(D_{\bar{N}}) = N_y(D_{\bar{N}})$, then $f^{\{x,y\}}(D_{\bar{N}}) = \{x, y\}$ and (b) if $N_x(D_{\bar{N}}) > N_y(D_{\bar{N}})$, then $f^{\{x,y\}}(D_{\bar{N}}) = \{x\}$.

(a) Suppose that given the electorate $\bar{N}$ and the preference profile $D_{\bar{N}} \in \mathcal{D}^{\bar{N}}$, $N_x(D_{\bar{N}}) = N_y(D_{\bar{N}})$ but the social choice function $f^{\{x,y\},\bar{N}}$ is such that $f^{\{x,y\}}(D_{\bar{N}}) = \{x\}$. Since the family $\left\{f^{\{x,y\},\bar{N}}\right\}_{\bar{N}}$ satisfies IIA by Lemma 1 and all social choice rules are anonymous by assumption, we deduce that for all electorates $\bar{N}$ and all preference profiles $D_{\bar{N}} \in \mathcal{D}^{\bar{N}}$, $f^{\{x,y\},\bar{N}}$ depends on the numbers $|N(D_{\bar{N}}; x, y)| = N_x(D_{\bar{N}}) + |\{i \in \bar{N} : x, y \in B(D_i)\}|$ and $|N(D_{\bar{N}}; y, x)| = N_y(D_{\bar{N}}) + |\{i \in \bar{N} : x, y \in B(D_i)\}|$. Let the permutation μ of K satisfy $\mu(x) = y$, $\mu(y) = x$ and $\mu(z) = z$ for all $z \in K \backslash \{x, y\}$. Neutrality implies that $f^{\{x,y\}}\left(\mu(D_{\bar{N}})\right) = \mu\left(f^{\{x,y\}}(D_{\bar{N}})\right) = \{y\}$. At this point it should be noted that $|N(D_{\bar{N}}; x, y)| = |N(D_{\bar{N}}; y, x)|$, because $N_x(D_{\bar{N}}) = N_y(D_{\bar{N}})$ by assumption. From this and the construction of μ we deduce that $|N(\mu\left(D_{\bar{N}}\right); x, y)| = |N(D_{\bar{N}}; x, y)|$ and $|N(\mu\left(D_{\bar{N}}\right); y, x)| = |N(D_{\bar{N}}; y, x)|$. Since the social social function $f^{\{x,y\},\bar{N}}$ just depends on these numbers, it has to be that $f^{\{x,y\},\bar{N}}(D_{\bar{N}}) = f^{\{x,y\},\bar{N}}\left(\mu(D_{\bar{N}})\right)$. We have reached a contradiction, because $f^{\{x,y\}}(D_{\bar{N}}) = \{x\}$ and $f^{\{x,y\}}\left(\mu(D_{\bar{N}})\right) = \{y\}$.

(b) Suppose that given the electorate $\bar{N}$ and the preference profile $D_{\bar{N}} \in \mathcal{D}^{\bar{N}}$, $N_x(D_{\bar{N}}) - N_y(D_{\bar{N}}) = 1$. Construct the preference profile $D'_{\bar{N}} \in \mathcal{D}^{\bar{N}}$ in the following way: For some individual i whose preference relation D_i is such that $x \in G(D_i)$ and $y \in B(D_i)$, the preference relation D'_i satisfies the condition $G(D'_i) = G(D_i) \backslash \{x\}$. Moreover, let $D'_j = D_j$ for all $j \neq i$. Since $N_x(D'_{\bar{N}}) = N_y(D'_{\bar{N}})$, it follows from part (a) that $f^{\{x,y\}}(D'_{\bar{N}}) = \{x, y\}$. Apply strict monotonicity to see that $f^{\{x,y\}}(D_{\bar{N}}) = \{x\}$.

Let $2 \leq m < \bar{n}$ and suppose that for all preference profiles $D_{\bar{N}} \in \mathcal{D}^{\bar{N}}$ which are such that $N_x(D_{\bar{N}}) - N_y(D_{\bar{N}}) = \bar{m} \leq m$, the condition $f^{\{x,y\}}(D_{\bar{N}}) = \{x\}$ holds. It remains to prove that if the preference profile $D_{\bar{N}} \in \mathcal{D}^{\bar{N}}$ satisfies $N_x(D_{\bar{N}}) - N_y(D_{\bar{N}}) = m + 1$, then $f^{\{x,y\}}(D_{\bar{N}}) = \{x\}$. Construct the preference profile $D'_{\bar{N}} \in \mathcal{D}^{\bar{N}}$ in the following way: For some individual i whose preference relation D_i is such that $x \in G(D_i)$ and $y \in B(D_i)$, the preference relation D'_i satisfies the condition $G(D'_i) = G(D_i) \backslash \{x\}$. Moreover, let $D'_j = D_j$ for all $j \neq i$. Since $N_x(D'_{\bar{N}}) - N_y(D'_{\bar{N}}) = \bar{m}$, it has to be that $f^{\{x,y\}}(D'_{\bar{N}}) = \{x\}$ by assumption. Apply strict monotonicity to see that $f^{\{x,y\}}(D_{\bar{N}}) = \{x\}$.

Finally, we show that the four properties imply Approval Voting; that is, for all sets of feasible alternatives $\bar{K}$, all electorates $\bar{N}$, and all preference profiles $D_{\bar{N}} \in \mathcal{D}^{\bar{N}}$, $x \in f^{\bar{K}}(D_{\bar{N}})$ if and only if $N_x(D_{\bar{N}}) \geq N_y(D_{\bar{N}})$ for all $y \in \bar{K}$. Suppose that $x \in f^{\bar{K}}(D_{\bar{N}})$. Then, $x \in f^{\{x,y\}}(D_{\bar{N}})$ for all $y \in \bar{K}\backslash\{x\}$ because the social choice rule $\{f^{\bar{K},\bar{N}}\}_{\bar{K},\bar{N}}$ is consistent in alternatives. This together with the fact that the family $\{f^{\{x,y\},\bar{N}}\}_{\bar{N}}$ orders x and y according to Approval Voting, implies that $N_x(D_{\bar{N}}) \geq N_y(D_{\bar{N}})$ for all $y \in \bar{K}$. To show the other inclusion suppose that $N_x(D_{\bar{N}}) \geq N_y(D_{\bar{N}})$ for all $y \in \bar{K}$. If there is an alternative $y \neq x$ such that $y \in f^{\bar{K}}(D_{\bar{N}})$, then $x \in f^{\bar{K}}(D_{\bar{N}})$ because $x \in f^{\{x,z\}}(D_{\bar{N}})$ for all $z \in \bar{K}\backslash\{x\}$ and $f^{\{x,y\}}(D_{\bar{N}}) = f^{\bar{K}}(D_{\bar{N}}) \cap \{x,y\}$ whenever the latter is non-empty. If there does not exist any alternative $y \neq x$ such that $y \in f^{\bar{K},\bar{N}}(D_{\bar{N}})$, then $x \in f^{\bar{K}}(D_{\bar{N}})$ because $f^{\bar{K}}(D_{\bar{N}}) \neq \emptyset$ by assumption. □

We show in the Appendix that Proposition 3.1 is tight. The result which is closest to Proposition 3.1 is due to Fishburn [35]. He characterizes Approval Voting as the only strategy-proof, neutral, anonymous and consistent family of social choice functions (it is allowed for a variable electorate whereas the set of feasible alternatives is assumed to be fixed). Using current notation consistency is defined as follows: Let $\bar{\mathcal{D}}$ be the domain of dichotomous preferences without the two preference relations indicating that an individual is indifferent between all alternatives. The family of social choice functions $\{f : \bar{\mathcal{D}}^{\bar{N}} \rightarrow 2^K\backslash\{\emptyset\}\}_{\bar{N}}$ is *consistent* if for all disjoint electorates $\hat{N}, \tilde{N}$, and all preference profiles $\bar{D}_{\hat{N}} \in \bar{\mathcal{D}}^{\hat{N}}$ and $\bar{\mathcal{D}}_{\tilde{N}} \in \bar{\mathcal{D}}^{\tilde{N}}$, $f(\bar{D}_{\hat{N}\cup\tilde{N}}) = f(\bar{D}_{\hat{N}}) \cap f(\bar{D}_{\tilde{N}})$ whenever $f(\bar{D}_{\hat{N}}) \cap f(\bar{D}_{\tilde{N}}) \neq \emptyset$. In the former definition, the preference profile $\bar{D}_{\hat{N}\cup\tilde{N}} \in \mathcal{D}^{\hat{N}\cup\tilde{N}}$ is obtained by unifying the other two preference profiles, e.g. $(\bar{D}_{\hat{N}\cup\tilde{N}})\big|_{\hat{N}} = \bar{D}_{\hat{N}}$ and $(\bar{D}_{\hat{N}\cup\tilde{N}})\big|_{\tilde{N}} = \bar{D}_{\tilde{N}}$. It is easy to see that the two characterizations are independent from each other, because neither does Fishburn's consistency condition imply strict monotonicity nor does strict monotonicity imply consistency.

Yet, only our characterization can be interpreted as an extension of the following version of May's Theorem [43]: Suppose that $k = 2$. The social choice function f :

$\mathcal{R}^N \to 2^K \backslash \{\emptyset\}$ is anonymous, neutral, and strictly monotone if and only if it the Majority (Condorcet) Rule; that is, for all $R \in \mathcal{R}^N$, $x \in f(R)$ if and only if $|\{i \in N : xR_iy\}| \geq |\{i \in N : yR_ix\}|$. Since Approval Voting is equivalent to the Condorcet Rule on the dichotomous preferences domain according to Brams and Fishburn [17], the main insight of Theorem 1 is that May's Theorem can be extended to any number of alternatives if we restrict our attention to social choice rules and add strategy-proofness to the original properties.

3.4 A Characterization with Efficiency

From now on we restrict our attention to the special case when all individuals reveal their preferences and all alternatives are feasible. The main idea of the second characterization is to explore the efficiency of Approval Voting. We use the standard notion of efficiency; that is, it is not possible to make some individual better off without hurting others.

Definition 3.6. The social choice function $f : \mathcal{D}^N \to 2^K \backslash \{\emptyset\}$ is said to be *efficient* if there does not exist a preference profile $D \in \mathcal{D}^N$ and a set $S \subseteq 2^K \backslash \{\emptyset\}$ such that $S \succsim_{D_i} f(D)$ for all $i \in N$ and $S \succ_{D_j} f(D)$ for some $j \in N$.

Approval Voting is not efficient for all reflexive, complete and transitive preference relations $\succsim_{D_i}$ on $2^K \backslash \{\emptyset\}$ that satisfy *condition* P and *condition* R, because if both good and bad alternatives are selected, then taking away a bad alternative can leave an individual indifferent.

Example 3.2. Suppose that $K = \{x, y, z\}$ and $n = 3$. If the preference profile $D \in \mathcal{D}^N$ is equal to $G(D_1) = \{x, y\}, G(D_2) = \{x, z\}$ and $G(D_3) = \{y, z\}$, then $f_A(D) = K$. Consider now the preference relations $\succsim_{D_1}$, $\succsim_{D_2}$ and $\succsim_{D_3}$ which are such that $\{x, y\} \succ_{D_1} K$, $\{x, y\} \sim_{D_2} K$ and $\{x, y\} \sim_{D_3} K$. These partial ordering neither contradict *condition* P nor *condition* R. Thus, Approval Voting is not efficient. ■

This problem disappears if we put more structure on the preference extension. In particular, we are going to assume from now on that every individual evaluates the set $S \subseteq 2^K \backslash \{\emptyset\}$ according to the proportion of good alternatives contained in S.

Definition 3.7. The preference relation $\succsim_{D_i}$ on $2^K \backslash \{\emptyset\}$ is said to be *cohesive with respect to* D_i whenever for all $S, T \in 2^K \backslash \{\emptyset\}$, $S \succsim_{D_i} T$ if and only if $\frac{|G(D_i) \cap S|}{|S|} \geq \frac{|G(D_i) \cap T|}{|T|}$ ($\succsim_{D_i}$ is strict whenever the inequality is strict).

Bogolmolnaia et al. [12] show that Approval Voting is efficient whenever preferences are cohesive (see their Proposition 1). Moreover, the following interpretation makes this preference extension particularly appealing: If we think of $f(D)$ as the set of pre-selected alternatives from which a unique winning alternative has to be determined via a lottery and individuals are expected utility maximizers, then individuals care only about the probability that a good alternative is chosen. If, in addition, individuals assign to all alternatives belonging to $f(D)$ the same winning probability, then the lottery with support on S is weakly preferred to the lottery with support on T if and only if $S \succsim_{D_i} T$.

Neutrality and strict symmetry are the other properties applied in the second characterization. The intuition of strict symmetry is simple: Suppose that there are two different preference profiles which differ from each other just because some alternative that is good for the first individual and bad for the second individual according to the first preference profile is good for the second individual and bad for the first individual according to the second preference profile. This variation in preferences should not provoke any change in the chosen set of alternatives. Formally, the preference profiles $D, D' \in \mathcal{D}^N$ are *x-symmetric* if for some pair of individuals (i, j), $G(D'_i) \cup \{x\} = G(D_i)$, $G(D'_j) = G(D_j) \cup \{x\}$, where $x \notin G(D'_i) \cup G(D_j)$, and $D'_l = D_l$ for all $l \neq i, j$.

Definition 3.8. The social choice function $f : \mathcal{D}^N \to 2^K \backslash \{\emptyset\}$ is said to be *strictly symmetric* if for all $x \in K$ and all x-symmetric preference profiles $D, D' \in \mathcal{D}^N$, $f(D) = f(D')$.

Lemma 2 shows that the social choice function $f : \mathcal{D}^N \to 2^K \backslash \{\emptyset\}$ depends on the k-dimensional vector $(N_x(D))_{x \in K}$ if and only if f is strictly symmetric. Formally, the social choice function $f : \mathcal{D}^N \to 2^K \backslash \{\emptyset\}$ *depends on the support of the alternatives* if for all preference profiles $D, D' \in \mathcal{D}^N$ which are such that $N_x(D) = N_x(D')$ for all $x \in K$, the condition $f(D) = f(D')$ holds.

Lemma 3.2. *The social choice function $f : \mathcal{D}^N \to 2^K \backslash \{\emptyset\}$ depends on the support of the alternatives if and only if it is strictly symmetric.*

Proof. It is easy to see that if f depends on the support of the alternatives, then f is strictly symmetric. To show the other inclusion suppose that $f : \mathcal{D}^N \to 2^K \backslash \{\emptyset\}$ is strictly symmetric and consider two preference profiles $D, D' \in \mathcal{D}^N$ which are such that $N_x(D) = N_x(D')$ for all $x \in K$. The following algorithm proves by double induction that $f(D) = f(D')$.

<u>Step 1:</u> In the beginning, define the set S_1 as $S_1 = K \backslash \{\{G(D_1) \cap G(D'_1)\} \cup \{B(D_1) \cap B(D'_1)\}\}$. Notice that S_1 consists of all alternatives which are differently ordered for individual 1 according to the preference relations D_1 and D'_1. Let $s_1 = |S_1| \geq 0$. If $s_1 = 0$, then $D_1 = D'_1$. In this case, let $D^1 \in \mathcal{D}^N$ be equal to $D \in \mathcal{D}^N$. We conclude that $f(D^1) = f(D)$ and $N_y(D^1) = N_y(D)$ for all $y \in K$. If $s_1 > 0$, then, without loss of generality, we can order the alternatives in S_1 according to the one-to-one mapping $g : S_1 \to \mathbb{N}_+$ which is such that $g(S_1) = \{1, 2, ..., s_1\}$ ($g(S_1)$ is the set obtained by applying the mapping g to all elements of S_1). Now proceed to step 1.1 of the algorithm.

<u>Step 1.1:</u> Suppose that $g(x) = 1$. If $x \in G(D_1)$ and $x \notin G(D'_1)$, then there is an individual $i > 1$ such that $x \in G(D'_i)$ and $x \notin G(D_i)$, because $N_y(D) = N_y(D')$ for all $y \in K$ and in particular for alternative $x \in S_1$. Let $1 < i \leq n$ be the smallest integer such that $x \in G(D'_i)$ and $x \notin G(D_i)$. Next, set the preference profile $D^{1.1}$ equal to $G(D_1^{1.1}) = G(D_1) \backslash \{x\}$, $G(D_i^{1.1}) = G(D_i) \cup \{x\}$ and $D_l^{1.1} = D_l$ for all $l \neq 1, i$. Since f is strictly symmetric, $f(D^{1.1}) = f(D)$. Notice that for all $y \in K$, $N_y(D^{1.1}) = N_y(D)$. On the other hand, if $x \in G(D'_1)$ and $x \notin G(D_1)$, then there is an individual $j > 1$ such that $x \in G(D_j)$ and $x \notin G(D'_j)$, because $N_y(D) = N_y(D')$ for all $y \in K$ and in particular for alternative $x \in S_1$. Let $1 < j \leq n$ be the smallest integer such that $x \in G(D_j)$ and $x \notin G(D'_j)$. Next, set the preference profile $D^{1.1}$ equal to $G(D_1^{1.1}) = G(D_1) \cup \{x\}$, $G(D_j^{1.1}) = G(D_j) \backslash \{x\}$

and $D_l^{1.1} = D_l$ for all $l \neq 1, j$. Since f is strictly symmetric, $f(D^{1.1}) = f(D)$. Notice that for all $y \in K$, $N_y(D^{1.1}) = N_y(D)$.

Let $M = \{1, ..., m\} \subseteq S_1$, $2 \leq m < s_1$, be the set of the first m alternatives of S_1. Suppose that for all $\bar{M} = \{1, ..., \bar{m}\} \subseteq M$, $f(D^{1.\bar{m}}) = f(D^{1.\bar{m}-1})$ and for all $y \in K$, $N_y(D^{1.\bar{m}}) = N_y(D^{1.\bar{m}-1})$. We show that given the set $M \cup \{m+1\}$, $f(D^{1.m+1}) = f(D^{1.m})$ and $N_y(D^{1.m+1}) = N_y(D^{1.m})$ for all $y \in K$.

Step 1.m+1: Suppose that $g(z) = m + 1$. If $z \in G(D_1^{1.m})$ and $z \notin G(D'_1)$, then there is an individual $i > 1$ such that $z \in G(D'_i)$ and $z \notin G(D_i^{1.m})$, because by the induction hypothesis $N_y(D^{1.m}) = N_y(D')$ for all $y \in K$ and in particular for alternative $z \in S_1$. Let $1 < i \leq n$ be the smallest integer such that $z \in G(D'_i)$ and $z \notin G(D_i^{1.m})$. Next, set the preference profile $D^{1.m+1}$ equal to $G(D_1^{1.m+1}) = G(D_1^{1.m}) \backslash \{z\}$, $G(D_i^{1.m+1}) = G(D_i^{1.m}) \cup \{z\}$ and $D_l^{1.m+1} = D_l^{1.m}$ for all $l \neq 1, i$. Since f is strictly symmetric, $f(D^{1.m+1}) = f(D^{1.m})$. Notice that for all $y \in K$, $N_y(D^{1.m+1}) = N_y(D^{1.m})$. On the other hand, if $z \in G(D'_1)$ and $z \notin G(D_1^{1.m})$, then there is an individual $j > 1$ such that $z \in G(D_j^{1.m})$ and $z \notin G(D'_j)$, because by the induction hypothesis $N_y(D^{1.m}) = N_y(D')$ for all $y \in K$ and in particular for alternative $z \in S_1$. Let $1 < j \leq n$ be the smallest integer such that $z \in G(D_j^{1.m})$ and $z \notin G(D'_j)$. Next, set the preference profile $D^{1.m+1}$ equal to $G(D_1^{1.m+1}) = G(D_1^{1.m}) \cup \{z\}$, $G(D_j^{1.m+1}) = G(D_j^{1.m}) \backslash \{z\}$ and $D_l^{1.m+1} = D_l^{1.m}$ for all $l \neq 1, j$. Since f is strictly symmetric, $f(D^{1.m+1}) = f(D^{1.m})$. Notice that for all $y \in K$, $N_y(D^{1.m+1}) = N_y(D^{1.m})$.

Let $D^1 \in \mathcal{D}^N$ be equal to $D^{1.s_1} \in \mathcal{D}^N$. So far it has been shown by induction that at $D^1 = (D'_1, D_2^1, ..., D_n^1) \in \mathcal{D}^N$, $f(D^1) = f(D)$ and $N_y(D^1) = N_y(D)$ for all $y \in K$. Let $2 \leq t < n - 1$, and, given the integer t, define the preference profile D^t as $D^t = \left(D'_1, ..., D'_t, D_{t+1}^t, ..., D_n^t\right) \in \mathcal{D}^N$. Suppose that for all $2 \leq \bar{t} \leq t$, $f\left(D^{\bar{t}}\right) = f\left(D^{\bar{t}-1}\right)$ and $N_y\left(D^{\bar{t}}\right) = N_y(D^{\bar{t}-1})$ for all $y \in K$. To finish the proof we have to show that $f(D^{t+1}) = f(D^t)$.

Step t+1: In the beginning, define the set S_{t+1} as $S_{t+1} = K \backslash \{\{G(D_{t+1}^t) \cap G(D'_{t+1})\} \cup \{B(D_{t+1}^t) \cap B(D'_{t+1})\}\}$. Notice that S_{t+1} consists of all alternatives which are ordered differently for individual $t + 1$ according to the preference relations D_{t+1}^t and D'_{t+1}. Let

$s_{t+1} = |S_{t+1}| \geq 0$. If $s_{t+1} = 0$, then $D^t_{t+1} = D'_{t+1}$. In this case, let $D^{t+1} \in \mathcal{D}^N$ be equal to $D^t \in \mathcal{D}^N$. We conclude that $f(D^{t+1}) = f(D^t)$ and $N_y(D^{t+1}) = N_y(D^t)$ for all $y \in K$. If $s_{t+1} > 0$, then, without loss of generality, we can order the alternatives in S_{t+1} according to the one-to-one mapping $g : S_{t+1} \to \mathbb{N}_+$ which is such that $g(S_{t+1}) = \{1, 2, ..., s_{t+1}\}$ ($g(S_{t+1})$ is the set obtained by applying the mapping g to all elements of S_{t+1}). Now proceed to step $t + 1.1$ of the algorithm.

Step t+1.1: Suppose that $g(x) = 1$. If $x \in G(D^t_{t+1})$ and $x \notin G(D'_{t+1})$, then there is an individual $i > t + 1$ such that $x \in G(D'_i)$ and $x \notin G(D^t_i)$, because by the induction hypothesis $N_y(D^t) = N_y(D')$ for all $y \in K$ (and in particular for alternative $x \in S_{t+1}$) and $D^t_l = D'_l$ for all $l \leq t$. Let $t + 1 < i \leq n$ be the smallest integer such that $x \in G(D'_i)$ and $x \notin G(D^t_i)$. Next, set the preference profile $D^{t+1.1}$ equal to $G(D^{t+1.1}_{t+1}) = G(D^t_{t+1})\backslash\{x\}$, $G(D^{t+1.1}_i) = G(D^t_i) \cup \{x\}$ and $D^{t+1.1}_l = D^t_l$ for all $l \neq t+1, i$. Since f is strictly symmetric, $f(D^{t+1.1}) = f(D^t)$. Notice that for all $y \in K$, $N_y(D^{t+1.1}) = N_y(D^t)$. On the other hand, if $x \in G(D'_{t+1})$ and $x \notin G(D^t_{t+1})$, then there exists an individual $j > t + 1$ such that $x \in G(D^t_j)$ and $x \notin G(D'_j)$, because by the induction hypothesis $N_y(D^t) = N_y(D')$ for all $y \in K$ (and in particular for alternative $x \in S_{t+1}$) and $D^t_l = D'_l$ for all $l \leq t$. Let $t + 1 < j \leq n$ be the smallest integer such that $x \in G(D^t_j)$ and $x \notin G(D'_j)$. Next, set the preference profile $D^{t+1.1}$ equal to $G(D^{t+1.1}_{t+1}) = G(D^t_{t+1}) \cup \{x\}$, $G(D^{t+1.1}_j) = G(D^t_j)\backslash\{x\}$ and $D^{t+1.1}_l = D^t_l$ for all $l \neq 1, j$. Since f is strictly symmetric, $f(D^{t+1.1}) = f(D^t)$. Notice that for all $y \in K$, $N_y(D^{t+1.1}) = N_y(D^t)$.

Let $M = \{1, ..., m\} \subseteq S_{t+1}$, $2 \leq m < s_{t+1}$, be the set of the first m alternatives of S_{t+1}. Suppose that for all $\bar{M} = \{1, ..., \bar{m}\} \subseteq M$, $f(D^{t+1.\bar{m}}) = f(D^{t+1.\bar{m}-1})$ and for all $y \in K$, $N_y(D^{t+1.\bar{m}}) = N_y(D^{t+1.\bar{m}-1})$. We show that given the set $M \cup \{m + 1\}$, $f(D^{t+1.m+1}) = f(D^{t+1.m})$ and $N_y(D^{t+1.m+1}) = N_y(D^{t+1.m})$ for all $y \in K$.

Step t+1.m+1: Suppose that $g(z) = m+1$. If $z \in G(D^{t+1.m}_1)$ and $z \notin G(D'_{t+1})$, then there is an individual $i > t+1$ such that $z \in G(D'_i)$ and $z \notin G(D^{t+1.m}_i)$, because by the induction hypothesis $N_y(D^{t+1.m}) = N_y(D')$ for all $y \in K$ (and in particular for alternative $z \in S_{t+1}$) and $D^t_l = D'_l$ for all $l \leq t$. Let $t + 1 < i \leq n$ be the smallest integer such that $z \in G(D'_i)$

and $z \notin G(D_i^{t+1.m})$. Next, set the preference profile $D^{t+1.m+1}$ equal to $G(D_1^{t+1.m+1}) = G(D_1^{t+1.m})\backslash\{z\}$, $G(D_i^{t+1.m+1}) = G(D_i^{t+1.m}) \cup \{z\}$ and $D_l^{t+1.m+1} = D_l^{t+1.m}$ for all $l \neq 1, i$. Since f is strictly symmetric, $f(D^{t+1.m+1}) = f(D^{t+1.m})$. Notice that for all $y \in K$, $N_y(D^{t+1.m+1}) = N_y(D^{t+1.m})$. On the other hand, if $z \in G(D'_{t+1})$ and $z \notin G(D_1^{t+1.m})$, then there is an individual $j > t+1$ such that $z \in G(D_j^{t+1.m})$ and $z \notin G(D'_j)$, because by the induction hypothesis $N_y(D^{t+1.m}) = N_y(D')$ for all $y \in K$ (and in particular for alternative $z \in S_{t+1}$) and $D_l^{t+1.m+1} = D_l^{t+1.m}$ for all $l \neq 1, i$. Let $t+1 < j \leq n$ be the smallest integer such that $z \in G(D_j^{t+1.m})$ and $z \notin G(D'_j)$. Next, set the preference profile $D^{t+1.m+1}$ equal to $G(D_1^{t+1.m+1}) = G(D_1^{t+1.m}) \cup \{z\}$, $G(D_j^{t+1.m+1}) = G(D_j^{t+1.m})\backslash\{z\}$ and $D_l^{t+1.m+1} = D_l^{t+1.m}$ for all $l \neq 1, j$. Since f is strictly symmetric, $f(D^{t+1.m+1}) = f(D^{t+1.m})$. Notice that for all $y \in K$, $N_y(D^{t+1.m+1}) = N_y(D^{t+1.m})$.

The algorithm finishes after $n-1$ steps, because the conditions $D_i^{n-1} = D'_i$ for all $i \neq n$ and $N_y(D^{n-1}) = N_y(D')$ for all $y \in K$ imply that $S_n = \emptyset$. □

At this point we are ready to state the second characterization.

Proposition 3.2. *The social choice function $f : \mathcal{D}^N \to 2^K\backslash\{\emptyset\}$ is strictly symmetric, neutral and efficient for the cohesive preference extension if and only if it is Approval Voting.*

Proof. Observe that Approval Voting is strictly symmetric, neutral and efficient. In order to prove the other inclusion suppose that the social choice function $f : \mathcal{D}^N \to 2^K\backslash\{\emptyset\}$ satisfies the three properties. It remains to show that for all preference profiles $D \in \mathcal{D}^N$, $x \in f(D)$ if and only if $N_x(D) \geq N_y(D)$ for all $y \in K$.

Consider the preference profile $D' \in \mathcal{D}^N$ which is such that for all $i \in N$ and $x \in K$, $x \in G(D'_i)$ if and only if $i \leq N_x(D)$. Since $N_y(D') = N_y(D)$ for all $y \in K$ and f depends on the support of the alternatives by Lemma 2, $f(D') = f(D)$. Define $p = \min_{y \in K} N_y(D')$ and $q = \max_{y \in K} N_y(D')$, respectively. We prove at first that $f(D') \subseteq G\left(D'_q\right)$ by efficiency. Suppose otherwise, that is for some $y \in K$, $y \in f(D')$ and $y \notin G\left(D'_q\right)$. By construction of the preference profile $D' \in \mathcal{D}^N$, $p \leq N_y(D') < q$. Let $x \in G\left(D'_q\right)$ and observe that for all $i \in N$, $\{x\} \succsim_{D_i} f(D')$ due to the cohesive preference extension. Moreover, for individual q, $\{x\} \succ_{D_q} f(D')$, because $x \in G\left(D'_q\right)$ and $y \notin G\left(D'_q\right)$. We conclude that f is not efficient, and therefore, $f(D') \subseteq G(D'_q)$.

Apply neutrality to obtain $f(D') = G\left(D'_q\right)$. Finally, we show that $x \in f(D)$ if and only if $N_x(D) \geq N_y(D)$ for all $y \in K$. Suppose that $x \in f(D)$. Then, $x \in G_q(D')$, because we have already seen that $f(D) = f(D') = G_q(D')$. Since $N_x(D') \geq N_y(D')$ for all $y \in K$ by construction and $N_z(D') = N_z(D)$ for all $z \in K$, the first inclusion has to be true. To show the other inclusion suppose that $N_x(D) \geq N_y(D)$ for all $y \in K$. Consider the preference profile $D' \in \mathcal{D}^N$ as described above. Since $N_z(D') = N_z(D)$ for all $z \in K$ by construction, $N_x(D') \geq N_y(D')$ for all $y \in K$. Thus, $x \in G_q(D')$ again by construction of $D' \in \mathcal{D}^N$. The result follows, because $f(D) = f(D') = G_q(D')$. □

We show in the Appendix that Proposition 3.2 is tight. Baigent and Xu [3] characterize the *choice aggregation procedure* corresponding to Approval Voting. Formally, let $M(R_i) \in 2^K$ be the alternatives individual i votes for when her/his preference relation is R_i. A choice aggregation procedure $c : \left(2^K\right)^N \to 2^K \backslash \{\emptyset\}$ aggregates the collective voting decisions $(M(R_1), ..., M(R_n))$ by selecting a non-empty set of alternatives. Choice aggregation procedures and social choice functions are not comparable for general preference domains, however they can be related to each other whenever preferences are dichotomous. To do so we interpret the alternatives an individual votes for as the set of her/his good alternatives (incentive-compatibility legitimates this approach); that is, for all $i \in N$ and all $D_i \in \mathcal{D}$, $M(D_i) = G(D_i)$. In this way, we are able to recover preferences from the observed voting decision, or, to say it differently, choice aggregation procedures and social choice functions coincide.

Baigent and Xu [3] characterize Approval Voting by means of Independence of Symmetric Substitutions (ISS), strict monotonicity and neutrality. Using our notation ISS is defined as follows: The social choice function $f : \mathcal{D}^N \to 2^K \backslash \{\emptyset\}$ satisfies *Independence of Symmetric Substitutions* (ISS) if for all preference profiles $D, D' \in \mathcal{D}^N$ which are such that for some pair of individuals (i, j) and some pair of alternatives (x, y), $G(D'_i) \cup \{y\} = G(D_i) \cup \{x\}$ and $G(D'_j) \cup \{x\} = G(D_j) \cup \{y\}$ where $y \notin G(D'_i) \cup G(D_j)$

and $x \notin G(D_i) \cup G(D'_j)$, and $D'_l = D_l$ for all $l \neq i, j$, the condition $f(D') = f(D)$ holds. Strict symmetry is obviously stronger than ISS, and therefore, it is a logical question to ask whether we can relax the latter property and characterize Approval Voting using ISS, neutrality and efficiency. In the Appendix, we provide a social social function different from Approval Voting that satisfies these properties. In this sense, the strengthening from ISS to strict symmetry is not only sufficient but also necessary in order to explore the efficiency of Approval Voting.

3.5 Conclusion

Brams and Fishburn [16] have recently outlined the practical importance of Apoproval Voting, and therefore, it is a logical consequence to ask for theoretical support of this aggregation rule. Our goal has been to look for new normative foundations of Approval Voting under the assumption of dichotomous preferences.

Working with dichotomous preferences is surely not innocuous, but if one wants to compare Approval Voting axiomatically with well known social choice functions such as Scoring Rules or Voting by Committees, then we necessarily have to restrict ourselves to this preference domain. To see this in an easy example suppose that there are three alternatives x, y and z and let the preference relation for individual i be such that xP_iyP_iz. In this case, individual i either votes for alternative x or for the set $\{x, y\}$ (see e.g. Luo et al. [42]). But if $M(R_i) = \{x, y\}$, then we cannot deduce that xP_iy. Similarly, if $M(R_i) = \{x\}$, then we do not know that yP_iz. Things become only different if preferences are dichotomous, because then individuals want to vote exactly for their set of good alternatives and the observed voting decision is as if individuals had fully revealed their preferences. Thus, the assumption of dichotomous preference is necessary if we want to define Approval Voting as a social choice function.

Finally, it should be noted that the literature on social choice has concentrated to large extend on the analysis of strict preferences although indifferences play an important role in a lot of problems of collective choice. The dichotomous preference domain constitutes without any doubt an important benchmark case in the analysis of those situations.

Appendix

Tightness of Proposition 3.1

We exhibit four social choice rules different from Approval Voting that violate only one different property each.

Strategy-Proofness:

The Borda Count is equivalent to Approval Voting on the dichotomous preference domain according to Proposition 2.1. Since all scoring rules are anonymous, neutral and strictly monotone, any scoring rule different from the Borda Count is manipulable on the dichotomous preference domain.

Neutrality:

Define the social choice rule $\left\{f_1^{\bar{K},\bar{N}} : \mathcal{D}^{\bar{N}} \rightarrow 2^K \backslash \{\emptyset\}\right\}_{\bar{K},\bar{N}}$ as follows: For all sets of feasible alternatives $\bar{K}$, all electorates $\bar{N}$, and all preference profiles $D_{\bar{N}} \in \mathcal{D}^{\bar{N}}$, $f_1^{\bar{K}}(D_{\bar{N}}) = \{x\}$ whenever $x \in f_A^{\bar{K}}(D_{\bar{N}})$. Otherwise, $f_1^{\bar{K}}(D_{\bar{N}}) = f_A^{\bar{K}}(D_{\bar{N}})$. Note that $\left\{f_1^{\bar{K},\bar{N}}\right\}_{\bar{K},\bar{N}}$ is anonymous, strictly monotone and strategy-proof. The following example illustrates that it is not neutral.

Let $K = \{x, y\}$ and $N = \{1, 2\}$. If the preference profile $D \in \mathcal{D}^N$ is equal to $G(D_1) = G(D_2) = \{x, y\}$, then $f_1^{\bar{K}}(D_{\bar{N}}) = \{x\}$. Define the permutation μ of K as $\mu(x) = y$ and $\mu(y) = x$. Since $\mu(D_{\bar{N}}) = D_{\bar{N}}$, it has to be that $f_1^{\bar{K}}(\mu(D_{\bar{N}})) = f_1^{\bar{K}}(D_{\bar{N}}) = \{x\}$. On the other hand, we observe that $\mu\left(f_1^{\bar{K}}(D_{\bar{N}})\right) = \{y\}$. Thus, $\mu\left(f_1^{\bar{K}}(D_{\bar{N}})\right) \neq f_1^{\bar{K}}(\mu(D_{\bar{N}}))$, a contradiction.

Anonymity:

Define the social choice rule $\left\{f_2^{\bar{K},\bar{N}} : \mathcal{D}^{\bar{N}} \rightarrow 2^{\bar{K}} \backslash \{\emptyset\}\right\}_{\bar{K},\bar{N}}$ as follows: For all sets of feasible alternatives $\bar{K}$, all electorates $\bar{N}$, and all preference profiles $D_{\bar{N}} \in \mathcal{D}^{\bar{N}}$, if the conditions

$\{1\} \in \bar{N}$, $0 < g(D_1) < \bar{k}$, and $G(D_1) \cap f_A^{\bar{K}}(D_{\bar{N}}) \neq \emptyset$ are satisfied, then $f_2^{\bar{K}}(D_{\bar{N}}) = G(D_1) \cap f_A^{\bar{K}}(D_{\bar{N}})$. Otherwise, $f_2^{\bar{K}}(D_{\bar{N}}) = f_A^{\bar{K}}(D_{\bar{N}})$. Note that the social choice rule $\left\{f_2^{\bar{K},\bar{N}}\right\}_{\bar{K},\bar{N}}$ is strategy-proof, strictly monotone and neutral. The following example illustrates that it is not anonymous.

Let $K = \{x, y\}$ and $N = \{1, 2\}$. Consider the preference profile $D \in \mathcal{D}^N$ corresponding to $G(D_1) = \{x\}$ and $G(D_2) = \{y\}$. In this case, $f_2(D) = \{x\}$. If the permutation σ of N is defined as $\sigma(1) = 2$ and $\sigma(2) = 1$, then $f_2\left(D_{\sigma(N)}\right) = \{y\}$. Thus $f_2\left(D_{\sigma(N)}\right) \neq f_2(D$, a contradiction.

Strict Monotonicity:

Given the set of feasible alternatives $\bar{K}$, the electorate $\bar{N}$, and the preference profile $D_{\bar{N}} \in \mathcal{D}^{\bar{N}}$, let $C \subseteq \bar{N}$ be such that $i \in C$ if and only if $0 < g(D_i) < \bar{k}$. Define the social choice rule $\left\{f_3^{\bar{K},\bar{N}} : \mathcal{D}^{\bar{N}} \to 2^{\bar{K}} \backslash \{\emptyset\}\right\}_{\bar{K},\bar{N}}$ as follows: For all sets of feasible alternatives $\bar{K}$, all electorates $\bar{N}$, and all preference profiles $D_{\bar{N}} \in \mathcal{D}^{\bar{N}}$, $f_3^{\bar{K}}(D_{\bar{N}}) = \{x \in \bar{K} : N_x(D_{\bar{N}}|_C) \geq 1\}$ whenever $C \neq \emptyset$. Otherwise, $f_3^{\bar{K}}(D_{\bar{N}}) = \bar{K}$. Observe that the social choice rule $\left\{f_3^{\bar{K},\bar{N}}\right\}_{\bar{K},\bar{N}}$ is strategy-proof, neutral and anonymous. The following example illustrates that it is not strictly monotone.

Suppose that $K = \{x, y, z\}$ and $N = \{1, 2\}$. Let the preference profiles $D, D' \in \mathcal{D}^N$ be such that $G(D_1) = \{y\}$, $G(D_1') = \{x, y\}$ and $G(D_2) = G(D_2') = \{x\}$. Then, $C = C' = \{1, 2\}$ which implies that $f_3^{\{x,y\}}(D) = f_3^{\{x,y\}}(D') = \{x, y\}$. This contradicts strict monotonicity.

Tightness of Proposition 3.2

We exhibit three social choice functions different from Approval Voting that violate only one different property each.

Neutrality:

The non-neutral social choice function f_1 is strictly symmetric and efficient.

Efficiency:

Define the social choice function $g : \mathcal{D}^N \to 2^K \backslash \{\emptyset\}$ as follows: For all preference profiles $D \in \mathcal{D}^N$, $g(D) = \{x \in K : N_x(D) \geq 1\}$ whenever this set is non-empty. Otherwise, $f(D) = K$. Note that g is strictly symmetric and neutral. The following example illustrates that it is not efficient.

Let $K = \{x, y, z\}$ and $N = \{1, 2, 3\}$. If preference profile $D \in \mathcal{D}^N$ is equal to $G(D_1) = \{x\}, G(D_2) = \{y\}$ and $G(D_3) = K$, then $g(D) = \{x, y, z\}$. Since $\{x, y\} \succ_{D_1} K$, $\{x, y\} \succ_{D_2} K$, and $\{x, y\} \sim_{D_3} K$ by the cohesive preference extension, we conclude that g is not efficient.

Strict Symmetry:

Define the social choice function $h : \mathcal{D}^N \to 2^K \backslash \{\emptyset\}$ as follows: If the preference profile $D \in \mathcal{D}^N$ is such that for some $i \in N$ and some $x \in K$, $G(D_i) = \{x\}$ and for all $j \neq i$, $G(D_j) = \{y\}$ for some $y \in K \backslash \{x\}$, then $h(D) = \{x, y\}$. Otherwise, $h(D) = f_A(D)$. Note that h is neutral, efficient and satisfies ISS. The following example illustrates that it is not strictly symmetric.

Let $K = \{x, y\}$ and $N = \{1, 2, 3\}$. If the preference profiles $D, D' \in \mathcal{D}^N$ are such that $G(D_1) = G(D'_1) = G(D_2) = \{x\}, G(D_3) = \{y\}$, $G(D'_2) = \{x, y\}$ and $G(D'_3) = \emptyset$, then $h(D) = \{x, y\}$ and $h(D') = \{x\}$. Since $h(D) \neq h(D')$, we conclude that h is not strictly symmetric.

Chapter 4

An Experimental Study of Truth-Telling in a Sender-Receiver Game

4.1 Introduction

In several situations individuals who lie about their private information can obtain a higher payoff at the costs of others.[1][2] But by behaving strategically individuals disrespect one of the oldest ethical principles, a social norm telling us *not to lie*. This tension between incentives and normative social behavior makes it difficult to predict the outcome of this type of interactions. It is our objective to show, with the help of an experiment, that in situations that can be modelled as a particularly simple *sender-receiver game*, a considerable number of subjects have preferences for truth-telling, whereas the rest of the subjects follow only material incentives.

Strategic information transmission, introduced by Crawford and Sobel [25], is an obvious way of modelling the tension described above. In this class of games, the "sender" has private information about the true state of the world. She transmits a message about

[1]This chapter is jointly written with Santiago Sánchez-Pagés from the University of Edinburgh.

[2]Examples include income tax evasion (Alingham and Sandmo [1]), oligopolistic competition (Galor [37]), financial advice (Morgan and Stocken [46]), and electoral competition (Heidhues and Lagerlof [40]).

the actual state to the "receiver" who takes a subsequent action that is payoff-relevant for both participants. The main insight of Crawford and Sobel [25] is that less information about the true state is transmitted as the preferences of the sender and the receiver become less aligned.

In the first experimental study on strategic information transmission, Dickhaut et. al [26] corroborated this theoretical prediction. More recently, Gneezy [38] has shown that if preferences are not aligned (whenever an outcome is good for the receiver it is bad for the sender and vice versa), then the probability of lying is increasing in the potential gains of the sender and decreasing in the potential loss of the receiver. Finally, Cai and Wang [20] have offered clear experimental evidence of an *overcommunication phenomenon*: Senders truthfully reveal their private information more often than predicted by the most informative equilibrium of the standard model of preference maximization. Although the authors explain this abnormality successfully by means of a behavioral type analysis (see among others Bosch-Domènech et. al [15], Costa-Gomes et. al [23] and Crawford [24]) and the Logit Quantal Response equilibrium concept (McKelvey and Palfrey [44] and [45]), they leave it as an open question whether the overcommunication phenomenon is caused by social preferences such as trust or honesty.

Our aim is to show that the tension between incentives and normative social behavior is the driving force underlying the overcommunication result. To this end, we study the experimental behavior of a group of subjects in two very similar constant-sum sender-receiver games. The *Benchmark Game* proceeds as follows: In the beginning of the game, one out of two payoff tables is randomly picked. The selected table determines players' (strictly positive) payoffs as a function of the receiver's action to be taken later on. Then, the sender, who is the only player informed about Nature's choice, submits a message about the actual payoff table; hence, she implicitly decides to tell the truth or to lie. After observing this message, the receiver takes an action that reveals whether he trusted or distrusted the sender. Finally, both participants are paid accordingly.

Since the payoff tables are constructed in such a way that the preferences of the sender and the receiver are not aligned, the sender does not have incentives to transmit information, or, to say it differently, the sender plays a strategy such that the posterior beliefs of the receiver remain equal to the prior beliefs. Given our model specification, only those strategies in which the sender lies with probability one-half generate these beliefs consistently and can thus be supported in equilibrium. In the first step of our analysis, we recover the overcommunication phenomenon: subjects playing the Benchmark Game in the role of the sender lie significantly less than predicted by the equilibrium analysis (Hypothesis 4.1).

To show that this result is caused by a considerable number of subjects with preferences for truth-telling we extend our original set-up. In the *Punishment Game*, the receiver is informed about the actual payoff table once he has taken an action. Then, he chooses between accepting the payoff distribution induced by the Benchmark Game and reducing the payoffs of both participants to zero.

According to the standard model of preference maximization individuals care only about their own payoffs, and therefore, the receiver should never punish the sender. But the limitations of the purely rational model are already well-documented. For example, the inequality aversion models of Fehr and Schmidt [33] and Bolton and Ockenfels [13] explain the empirical observation that some individuals are willing to pay money in order to reduce income disparities. Recently Brandts and Charness [19] have found evidence of even more complex preferences. Individuals do not only take into account the whole payoff distribution, rather the notion of *procedural justice*[3] - the utility attached to a payoff distribution depends on how this distribution has been reached - plays a crucial role in socio-economic interactions. As a matter of fact, the authors study in the laboratory a game in which the sender transmits a message regarding her/his intended play in a 2×2 simultaneous move game and show that the receiver's willingness to punish the sender after revealing the result from the simultaneous move game depends on whether or not the sender played according to the reported message.

We derive our predictions with respect to the extended game by looking at the punish-

[3] The concept of procedural justice has been introduced in decision theory by Sen [55] as an extension of the standard model of preference maximization over material outcomes.

ment rates after different histories. Since the game is symmetric, histories can be summarized by whether or not a message is truthful and whether or not the receiver trusts that message. The punishment rates are equal to 0% after history (truth,trust), 1.6% after (lie,distrust), 5.4% after (truth,distrust), and 25.2% after (lie,trust). The fact that the payoff distribution after history (truth,distrust) is equal to the one after (lie,trust) allows us then to confirm the importance of procedural justice in socio-economic interactions (Hypothesis 4.2).

Finally, we show that the overcommunication phenomenon can be explained in terms of social preferences for truth-telling. First, we identify all subjects with strong concerns for procedural justice. In particular, we find find that 15 out of 66 individuals punish liars frequently after the history (lie,trust). Not surprisingly this group of individuals accounts for 90% of all punishments after this history. Then, we analyze how these subjects behave in the role of the sender. It turns out that they tell the truth in over 70% of all observations whereas the rest of subjects do so only in 52% (the overall percentage of truth-telling in the Punishment Game is equal to 57%). This result supports our intuition that individuals with a strong sense for procedural justice should, consistently, be responsible for the overcommunication result (Hypothesis 4.3).

We proceed as follows: In the next Section, we formally introduce the games and our experimental hypotheses. In Section 3, we explain the experimental procedures. In the following Section, we present our results. We conclude in Section 5. The proof of Proposition 4.1 and the instructions of the Punishment Game can be found in the Appendix.

4.2 Theoretic Analysis and Experimental Predictions

In this Section, we introduce the Benchmark and the Punishment Game and derive several null hypotheses from the corresponding sequential equilibria. Moreover, we present our alternative hypotheses deduced from the overcommunication phenomenon and the incorporation of procedural justice.

The Benchmark Game

Let $N = \{\text{sender,receiver}\}$ be the set of players. At the beginning of the game, Nature picks payoff table A and B with equal probability, e.g. $p(A) = p(B) = 0.5$. Only the sender is informed about the payoff table actually chosen. Selecting table $x \in \{A, B\}$ means that the final payoffs are realized according to x. Both tables depend only on the action U or D taken by the receiver later on.

Table A	Sender	Receiver
Action U	2	1
Action D	1	2

Table B	Sender	Receiver
Action U	1	2
Action D	2	1

Table 4.1: Payoff Tables

After the sender has been informed, she chooses a mixed strategy with support on the message space $M = \{A, B\}$. Formally, if Nature selects table A, the sender communicates with probability p_A that table A represents the actual payoff scheme. Thus, she lies in this case with probability $1 - p_A$. Similarly, if Nature selects table B, then the sender communicates with probability $1 - p_B$ that table B represents the actual payoff scheme. Thus, she lies in this case with probability p_B.

Next, we describe the receiver's belief system. If $m = \{A\}$ (the sender transmits message A), the receiver believes with probability $\mu(A|A)$ that the actual payoff scheme is represented by table A whereas he thinks with probability $\mu(B|A) = 1 - \mu(A|A)$ that table B is the one determining payoffs. If $m = \{B\}$, the receiver believes with probability $\eta(A|B)$ that table A determines payoffs and with probability $\eta(B|B) = 1 - \eta(A|B)$ that table B is the one doing so. Taking into account these beliefs, the receiver chooses a mixed strategy with support on the action set $\mathcal{A} = \{U, D\}$. Formally, if $m = \{A\}$, the receiver takes action U with probability q_A and action D with probability $1 - q_A$. Similarly, if $m = \{B\}$, the receiver takes action U with probability q_B and action D with probability $1 - q_B$. Finally, both individuals receive their payments. ■

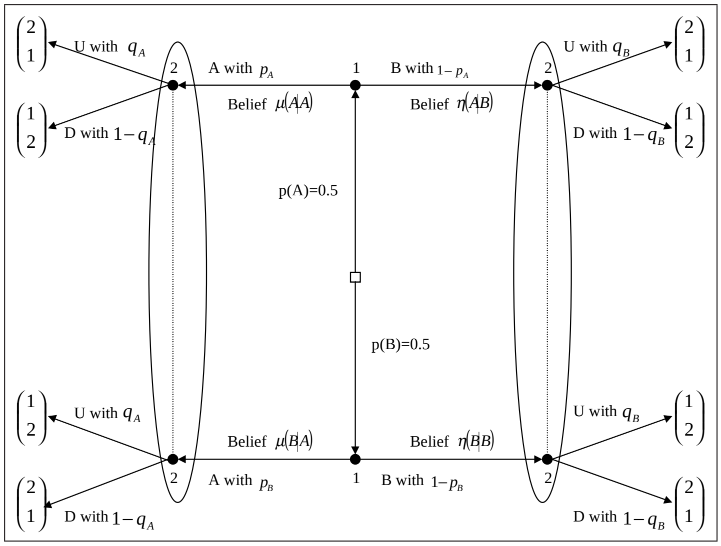


Figure 4.1: The Benchmark Game

The Benchmark Game is well suited to analyze the tension between social preferences for truth-telling and material incentives for two reasons. First, in order to minimize the possibility of mistakes made by subjects, the Benchmark Game has a simple payoff structure and a very intuitive set of equilibria. Second, truth-telling is a dichotomous choice in the sense that (a) there are only two state variables and (b) the sender's strategy set boils down to the messages *truth* and *lie*. This is important, because otherwise a message may contain a richer meaning. To see this suppose that the state and message space are both equal to $\{1, 2, 3\}$. In this case individuals do not only tell the truth or lie, because they also choose a "level" of deceit whenever the true state is $\{1\}$ or $\{3\}$. Hence, richer state and message spaces give room to a wide variety of behaviors and to a complexity that lies out of the scope of the paper.[4]

Proposition 4.1. *The set of sequential equilibria of the Benchmark Game is given by the set of strategies* $(p_A^*, p_B^*, q_A^*, q_B^*) = (p, p, q, q)$*, where* $p, q \in [0, 1]$*, and the supporting belief system* $(\mu^*(A|A), \eta^*(B|B)) = \left(\frac{1}{2}, \frac{1}{2}\right)$.

[4]The importance of the size of the message space is reported by Blume et. al [11]. The authors show that in a sender-receiver game with multiple equilibria it depends on the size of message space whether subjects converge to play a separating or a pooling equilibrium.

Proof. See the Appendix. □

The intuition of Proposition 4.1 is as follows: Since the preferences of the sender and the receiver are not aligned, the sender plays a strategy that leaves the receiver's prior beliefs unchanged. The strategies generating these posterior beliefs in a consistent manner are all those in which the sender submits message A with a constant probability, e.g. $p_A = p_B = p \in [0, 1]$. To see this note that if the sender plays for example the strategy "always transmit message A" (this strategy is equal to $p_A = p_B = 1$), then the receiver does not get any additional information from the message. Hence, the receiver can as well ignore it. This game becomes thus equivalent to the following one: Nature selects the tables A and B with equal probability and the receiver chooses q (the probability to play U) to maximize his expected payoff. Since the expected payoff is equal to $p(A)(q + 2(1-q)) + p(B)(2q + 1 - q) = 1.5$ and thus independent of q, any constant strategy $q_A = q_B = q \in [0, 1]$ is optimal.

The set of pooling equilibria is quite large, yet there is an easy way to identify them. Given p_A and p_B, the probability that the sender lies in the Benchmark Game is equal to $l_1(p_A, p_B) = p(A)(1 - p_A) + p(B)p_B$. With a slight abuse of notation let $p(m = x)$ be the probability that the sender transmits message $x \in \{A, B\}$ given p_A and p_B. Then, the probability that the receiver trusts the sender in the Benchmark Game is equal to $t_1(p_A, p_B) = p(m = A)\mu(A|A) + p(m = B)\eta(B|B)$. Proposition 2 establishes that the sender lies with probability one-half in any sequential equilibrium of the Benchmark Game, a strategy foreseen correctly by the receiver in terms of trust.

Proposition 4.2. *Let (p_A^*, p_B^*) be an equilibrium strategy for the sender in the Benchmark Game. Then, $l_1(p_A^*, p_B^*) = t_1(p_A^*, p_B^*) = \frac{1}{2}$.*

Proof. The proof is straightforward and thus omitted. □

Our first null hypothesis is given by Proposition 4.2. The corresponding alternative hypothesis is divided into two parts. First, the sender should lie less than predicted by the standard model according to the overcommunication phenomenon of Cai and Wang [20].

If this prediction is true, the best response function of the receiver dictates to take action D after message A and action U after message B, or, to say it differently, the receiver should always take the action *trust*.[5] This observation leads us to hypothesize additionally that the receivers take the action *trust* in more than fifty percent of all occasions.

HYPOTHESIS 4.1. *In the Benchmark Game, the senders lie in less than fifty percent and the receivers trust the senders in more than fifty percent of all observations.*

Next, we introduce punishments into the Benchmark Game in order to study whether the source of the overcommunication phenomenon are preferences for truth-telling shared by a considerable number of subjects.

The Punishment Game

The Punishment Game extends the Benchmark Game. Let H be the set of all histories of the Benchmark Game. Given $h \in H$, with probability $r_2(h) \in [0, 1]$ the receiver reduces the payoffs of both participants to 0 and with probability $1 - r_2(h)$ he accepts the payoff distribution induced by the Benchmark Game. Finally, both individuals receive their payments. ■

Given the strategy (p_A, p_B) of the sender in the Punishment Game, $l_2(p_A, p_B)$ and $t_2(p_A, p_B)$ denote the probabilities that the sender lies and that the receiver trusts the sender's message, respectively. It is easy to calculate the set of sequential equilibria of the Punishment Game, because from a purely materialistic point of view it is never optimal for the receiver to reduce payoffs. This observation allows us to draw the following conclusion.

Proposition 4.3. *In all sequential equilibria of the Punishment Game, (a) for all* $h \in H, r_2^*(h) = 0$ *and (b)* $l_2(p_A^*, p_B^*) = t_2(p_A^*, p_B^*) = \frac{1}{2}$.

Proof. The proof is straightforward and thus omitted. □

[5] We interpret the revealed decision of the receiver as the result from a maximization process involving subjective beliefs about the truthfulness of the message (see for example Farrell and Rabin [32]).

The second null hypothesis is given by Proposition 4.3. To derive the corresponding alternative hypothesis note that the set H can be summarized by the histories h_1 = (truth,trust), h_2 = (truth,distrust), h_3 = (lie,trust), and h_4 = (lie,distrust).[6] Different punishment rates after these histories will reveal the presence of social preferences.

The inequality aversion models of Fehr and Schmidt [33] and Bolton and Ockenfels [13] take into account that some individuals care not only about their own payoff but about the whole payoff distribution. Brandts and Charness [19] have shown that receivers punish the senders more often if a payoff distribution results from a deceptive message, and therefore, the utility attached to a particular payoff distribution also depends on how it has been reached. Following this experimental finding, the receiver should punish the sender more frequently after history h_3 = (lie,trust) than after history h_2 = (truth,distrust) although both payoff distributions are identical. Still, we expect the punishment rate after history h_2 = (truth,distrust) to be strictly positive, because some individuals may be inequality averse and prefer the payoff distribution (0,0) over (2,1). We do not expect any punishments after the histories h_1 = (truth,trust) and h_4 = (lie,distrust), because the receiver interpreted the message correctly and the resulting payoff distribution (1,2) is favorable to him. If we reject the null hypothesis in favor of the alternative one, a sender who lies is in more danger of being punished than a truth-teller. As a consequence, we hypothesize that truth-telling is enhanced in the Punishment Game with respect to the Benchmark Game and that the receivers trust more in the former than in the latter.

HYPOTHESIS 4.2. *In the Punishment Game, the receivers punish the senders only after the histories $h_2 = (truth, distrust)$ and $h_3 = (lie, trust)$ with the punishment rate being higher after h_3. Moreover, the senders lie less and the receivers trust more in the Punishment than in the Benchmark Game.*

According to our main hypothesis the overcommunication phenomenon is caused by number of individuals with preferences for truth-telling. To check this we perform a final

[6] In our experimental sessions we do not ask subjects to elicit their mixed strategies, rather we derive them from the repeated observation of pure strategies. Then, since the payoff tables A and B are symmetric and the probabilities $p(A)$ and $p(B)$ are identical, we can write the set H as described.

consistency test on the Punishment Game. After observing the experimental results, we divide our subject pool into two different groups, one group of subjects punishing liars frequently after history h_3 = (lie,trust) and another group containing the rest of the subjects. Given this division, the third null hypothesis states that there is no difference in the level of truth-telling between the two groups when subjects play the Punishment Game as senders.[7] The corresponding alternative hypothesis, on the other hand, states that the group of subjects with a high sense of procedural justice accounts for most of the overcommunication phenomenon; that is, these subjects tell the truth very often whereas the rest of the subjects lie in about fifty percent of the occasions.

HYPOTHESIS 4.3. *The group of subjects punishing liars frequently after history h_3 = (lie,trust) accounts for most of the overcommunication phenomenon.*

4.3 Experimental Design and Procedures

We conducted our experimental sessions at the University of Edinburgh in May 2004. Since all economic students at this university have an E-mail account associated to their matriculation number, we promoted the experiment mainly via electronic newsletters. Students from other academic disciplines were recruited through flyers distributed on the campus and further announcements made on information boards. As a result, 132 undergraduate students from nearly all faculties participated in one of our experimental sessions. We organized a total of ten sessions, five on the Benchmark and five on the Punishment Game. Twelve subjects participated in the first four sessions and eighteen subjects in the fifth and last session of each treatment. No subject took part in more than one session.

To perform the experiment we employed the computer software Z-Tree developed by Fischbacher [34]. At the beginning of a session, subjects met in a computer room and sat down in front of one of the computers. The computers were placed in such a way that all

[7] We use a role rotation mechanism in our experimental sessions so that every subject plays the Punishment Game half of the time in each role. For more on this see the next section.

subjects could only look at their own screen. We placed next to each computer a closed envelope containing instructions, a questionnaire, and a payment receipt. After subjects had filled out the questionnaire we read the instructions aloud (see the Appendix for the instructions corresponding to the Punishment Game).

Before the first round of a session, the computer randomly divided subjects into groups of six without revealing the matching. We informed every subject that s/he would only play against subjects belonging to the same group. Therefore, the fact that the number of subjects differed across sessions should not matter. So we implicitly divided our subject pool into a total of twenty-two groups of six subjects, eleven groups playing each treatment. In each of the fifty rounds of an experimental session the computer matched the subjects belonging to the same group into three new pairs and assigned different roles (sender or receiver) within pairs. The matchings were balanced so that after fifty rounds every subject played the game exactly ten times against each of her/his five opponents. Moreover, every subject met every opponent five times in each role.

In every round, after pairs had been formed and roles had been assigned, the sender was informed of whether table A or B had been selected. Then, the sender transmitted a message from the message space $M = \{A, B\}$ telling the receiver which table corresponds to the actual payoff scheme. Afterwards, the receiver chose an action from the action set $\mathcal{A} = \{U, D\}$. This constituted the end of the round in the Benchmark Game. In the sessions corresponding to the Punishment Game, the receiver was further informed about the induced payoffs of her/his action. Finally, s/he had to decide between accepting these payoffs or reducing the payoff of both participants to zero.

At the end of a session, we called subjects one by one to step forward to the control desk for payment. In addition to the five pounds show up fee, subjects received ten pence per point obtained. As a result, the average payment in the one hour session corresponding to the Benchmark and the Punishment Game was equal to 12.5 pounds and 11.74 pounds, respectively.

4.4 Results

4.4.1 Overcommunication in the Benchmark Game

According to our first null hypothesis, the senders lie in the Benchmark Game with probability one-half. In the histogram in the left part of Figure 4.2, we represent the frequencies of truthful messages in the sessions corresponding to the Benchmark Game. Since a subject was exactly 25 times in the role of the sender, in equilibrium s/he should tell the truth 12.5 times. The data seem to be slightly shifted to the right of the theoretical mean, but it is not clear-cut enough to reject the null hypothesis immediately. In the right panel of Figure 4.2, we observe that the percentage of subjects telling the truth is extraordinary high in the first rounds and declines over time in such a way that it stays on average just above the 50%-line predicted by the standard theory. We eliminate this learning effect by excluding the data from the first ten rounds in our statistical analysis.

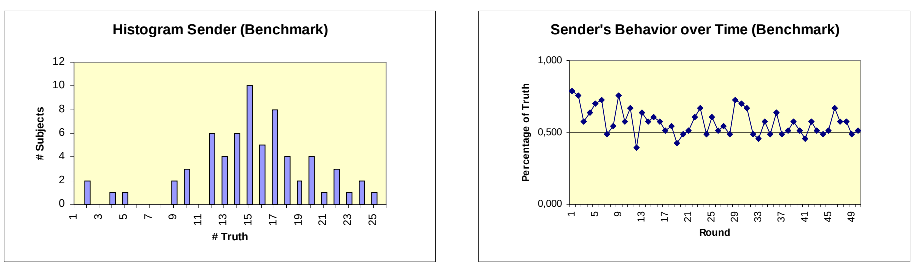


Figure 4.2: Sender's Behavior in the Benchmark Game

Since subjects belonging to the same group play the Benchmark Game more than once against the other group members, actions within a given group are likely to be correlated over time. One way of obtaining independent observations is to calculate for every group the percentage of truthful messages over the last forty rounds. This procedure allows us to derive a total of eleven independent observations, one for each group. The overall percentage of truth-telling in the last forty rounds is equal to 55.07%, a percentage significantly greater than 50% (p-value of the one-tailed Wilcoxon rank-sum test $= 0.0615$; p-value of the one-tailed t-test $= 0.0459$).

Next, we provide evidence in favor of the second part of Hypothesis 4.1, namely that the receivers adjust their beliefs in the correct direction and trust the senders in more than fifty percent of all occasions. To this end, we interpret the action of the receiver as the result of a maximization process involving subjective beliefs about the truthfulness of the message. For example, if a subject observes message A and takes action D afterwards, then this action reveals in our understanding that the subject trusted the sender's message. In the histogram in the left panel of Figure 4.3 we can clearly see that a lot of receivers trust more often than the theoretical prediction of 12.5 times. Moreover, the evolution of this percentage over time (the right panel of Figure 4.3) is such that it is particularly low in the first rounds before it stabilizes well above the 50%-line.

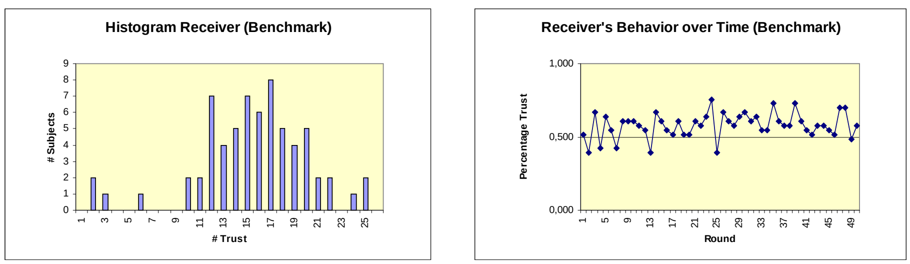


Figure 4.3: Receiver's Behavior in the Benchmark Game

In the last forty rounds of the experiment the receivers trusted the senders' messages in 58.7% of all observations. This value is significantly greater than the theoretical prediction. (p-value of the one-tailed Wilcoxon rank-sum test < 0.0001; p-value of the one-tailed t-test < 0.0001). Hence, we reject the prediction of the standard model in favor of Hypothesis 4.1.

4.4.2 Procedural Justice in the Punishment Game

So far we have shown the existence of an overcommunication phenomenon in the Benchmark Game. Analyzing the Punishment Game will help us to identify the origin of this result. In Table 4.2 below we present the punishment behavior of the receivers. For consistency reasons we only consider punishments in the last forty rounds which amount to a total of 1320 observations (11 groups of 40 rounds and 3 observations per round). The senders told the truth 740 times and lied in 580 occasions. The receivers trusted 520 times when the sender had told the truth and 396 times when the sender had lied.

	Truth	Lie
Trust	0	0,25253
Distrust	0,05455	0,0163

Table 4.2: Punishment Behavior

The punishment rate is highest, more than 25%, after history h_3 = (lie,trust). We use the normal approximation of the binomial distribution in order to establish that this proportion is significantly greater than zero (p-value of the one-tailed Z-test < 0.0001). We also find, as expected, that the punishment rate after history h_2 = (truth,distrust) is significantly greater than zero. We attribute the positive punishment rate after history h_4 = (lie,distrust) to mistakes made by some subjects. Yet, our main prediction is confirmed: The willingness to punish the sender depends on whether or not a payoff distribution has been reached by means of a deceptive message, because the punishment rate after history h_3 = (lie,trust) is greater than the one after history h_2 = (truth,distrust). A test of equal proportions confirms this observation (p-value of the one-tailed Z-test < 0.0001).

We investigate next whether subjects behave consistently across the two treatments. The histogram in left panel of Figure 4.4 looks quite similar to the one corresponding to sender's behavior in the Benchmark Game although it seems that the shift to the right from the theoretical mean has increased. In the right panel of Figure 4.4, we observe that the percentage of subjects telling the truth is quite high in the first rounds and declines over time, a behavior we have already encountered before. Nevertheless, in the latter rounds there are now less values below the fifty percent line.

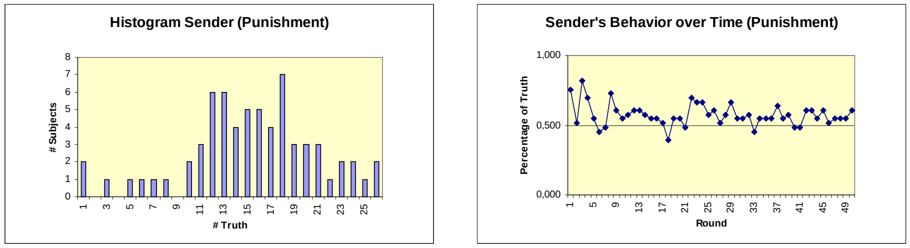


Figure 4.4: Sender's Behavior in the Punishment Game

The percentage of subjects telling the truth in the last forty rounds of the Punishment Game is equal to 56.29%. This percentage is significantly greater than the equilibrium prediction (p-value of the one-tailed Wilcoxon rank-sum test = 0.0499; p-value of the one-tailed t-test = 0.0343), but it is not significantly greater than the corresponding value for the Benchmark Game (p-value of the one-tailed Wilcoxon rank-sum test = 0.40; p-value of the one-tailed t-test = 0.385).

The picture looks quite different if we compare the receivers' behavior across the two treatments. The histogram in the left panel of Figure 4.5 indicates that the receivers trust more in the Punishment than in the Benchmark Game. This intuition is confirmed in the right panel of Figure 4.5, because the percentage of receivers trusting the sender seems to increase over time and stays well above the equilibrium prediction. On the aggregate, the percentage of trustful receivers in the last forty rounds is equal to 69.3%. This percentage is significantly greater than the corresponding value of the Benchmark Game (p-value of the one-tailed Wilcoxon rank-sum and t-test < 0.0001).

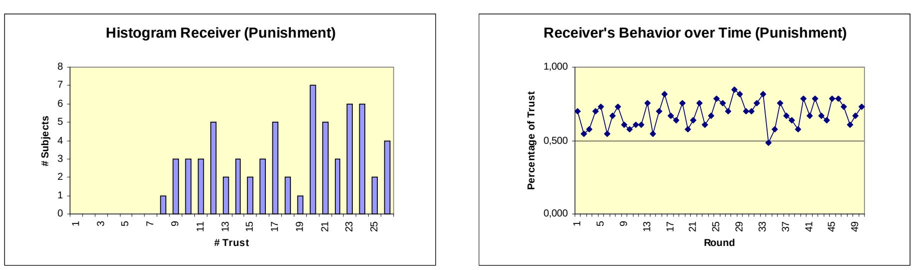


Figure 4.5: Receiver's Behavior in the Punishment Game

To summarize: We have confirmed the importance of procedural justice in socio-economic interactions. Moreover, the introduction of punishments seems to induce the receivers to believe that the senders will often tell the truth in order to avoid a possible moral outrage caused by deceptive messages. But on the contrary, when subjects play as senders they seem to consider the punishment as an incredible threat, because barely change their behavior with respect to the original set-up.

4.4.3 The Separability Hypothesis

In the previous two parts of this Section we have prepared the ground for our main contribution, namely that that the tension between normative social behavior and incentives is driving the overcommunication phenomenon. After observing the experimental results, we divide our subject pool into two groups, one group containing all those subjects who punish liars frequently after history $h_3 =$ (lie,trust) and another group containing the rest of subjects. We obtain this division in the following way: In the last forty rounds of an experimental session corresponding to the Punishment Game every subject is twenty times in the role of the receiver. Since the sender lies with probability 0.437 and the receiver trusts the message with probability 0.694, every subject plays, in expected terms, the history $h_3 =$ (lie,trust) 6.06 times in the role of the receiver. The punishment rate after h_3 is equal to 0.2525, and therefore, every subject is expected to punish the sender 1.53 times. Hence, all subjects that punish the sender in at least three occasions after h_3 reveal serious concerns for procedural justice. This condition is met by fifteen out of sixty-six subjects. Not surprisingly, this group of subjects accounts for 90% of all punishments.

Given this classification, the role rotation mechanism allows us to study how these fifteen subjects behave in the Punishment Game. On the aggregate, they tell the truth in 70.66% of all observations. This probability is significantly greater than 56.29%, the percentage of truth-telling corresponding to the whole subject pool (p-value of the one-tailed Z-test < 0.0001). The rest of the subjects, on the other hand, tell the truth in only 52.05% of the cases, a percentage not significantly greater than the equilibrium prediction

(p-value of the one-tailed Z-test ≥ 0.0945). Therefore, we reject the third null hypothesis - the percentage is the same for both groups of subjects - in favor of Hypothesis 4.3. This result suggests not only that individuals with a strong notion of procedural justice behave consistently across roles, rather they are responsible for nearly all information transmitted by the senders. This interpretation is further strengthened if we analyze how the beliefs of these two groups vary. Subjects with a serious notion for procedural justice trust the senders' message in 86% of all occasions, whereas the rest of subjects do so only in 64,51%.

4.5 Conclusion

Communication is the most natural way how to exchange information. Experimental studies such as the ones of Duffy and Feltovich [27] and [28] have shown that individuals are able to achieve Pareto improving allocations by means of cheap talk. In particular, the authors show that if subjects announce to cooperate in the Prisoner's Dilemma, then this message often reflects the truth. Moreover, receivers reciprocate and cooperate as well so the Pareto-efficient outcome is sometimes implemented. On the other hand, Crawford [24] shows that in some sequential equilibria of a sender-receiver game a rational individual can feint a boundedly rational one. These results raise some questions. In which situations can the receiver trust the senders' messages? And why do the senders transmit truthful messages if incentives suggest otherwise? Our aim was to show that the overcommunication phenomenon is not necessarily due to a lack of sophistication or rationality but results from the fact that some individuals take into account social norms such as truth-telling.

To this end, we studied the behavior of a group of subjects in a simple game of strategic information transmission. In the first step, we recovered the overcommunication phenomenon of Cai and Wang [20] (e.g. on the aggregate the senders tell the truth more often than predicted by the standard model of preference maximization). Then, we introduced punishments and showed that, in accordance with the results of Brandts and

Charness [19], the willingness to punish the sender is higher after a deceptive message. Finally, we sustained our main hypothesis by showing that if we subtract from our subject pool the group of subjects who punish liars frequently after a deceptive message, then that very same group tells the truth very often whereas the rest of the subjects behave roughly according to the standard equilibrium prediction. Thus, if moral subjects are excluded, the overcommunication phenomenon vanishes.

The existence of moral individuals who reject material incentives to misbehave opens some fascinating questions: What are the implications on mechanism design? Or on the organization of the firm? And on the elaboration of policy prescriptions?

Appendix

Proof of Proposition 4.1

Recall that p_A (or p_B, respectively) denotes the probability that the sender submits message A when the actual payoff scheme is represented by table A (or table B, respectively). We divide our analysis into three different cases.

Case 1: Suppose that $0 < p_A + p_B < 2$. We derive the best response correspondence for the receiver who takes the strategy (p_A, p_B) of the sender as given. Suppose that the sender transmits message A. By sequential rationality the receiver updates his beliefs according to Bayes' rule, and therefore, he thinks that the probability $\mu(\theta = A|m = A)$ (e.g. the true payoff scheme is given by table A conditional on message A) is equal to

$$\mu(\theta = A|m = A) = \frac{p(m=A|\theta=A)p(A)}{p(m=A)} = \frac{0.5p_A}{0.5p_A+0.5p_B} = \frac{p_A}{p_A+p_B} \equiv \mu.$$

Let $\mu(\theta = B|m = A) = 1 - \mu$ be the belief that table B represents payoffs when the sender submitted message A before. Given μ, the receiver chooses q_A (the probability to take action U conditional on message A) in order to

$$\max_{q_A} \quad (\mu (q_A + 2 (1 - q_A)) + (1 - \mu) (2q_A + 1 - q_A)) .$$

This maximization problem is equivalent to

$$\max_{q_A} \quad (1 + \mu + q_A (1 - 2\mu)),$$

and therefore, the best response correspondence for the receiver is

$$q_A^*(\mu) = \begin{cases} 1 & \text{if } \mu < \frac{1}{2} \\ [0,1] & \text{if } \mu = \frac{1}{2} \\ 0 & \text{if } \mu > \frac{1}{2}, \end{cases} \quad \text{, or,} \quad q_A^*(p_A, p_B) = \begin{cases} 1 & \text{if } p_A < p_B \\ [0,1] & \text{if } p_A = p_B \\ 0 & \text{if } p_A > p_B. \end{cases}$$

If, on the other hand, the sender submits message B, then the belief that the actual payoff scheme is represented by table B is equal to

$$\eta(\theta = B|m = B) = \frac{p(m=B|\theta=B)p(B)}{p(m=B)} = \frac{0.5(1-p_B)}{0.5(1-p_A)+0.5(1-p_B)} = \frac{1-p_B}{2-p_A-p_B} \equiv \eta.$$

Let $\eta(\theta = A|m = B) = 1 - \eta$ be the belief that table A represents the payoff scheme when the sender submitted message B before. Given η, the receiver chooses q_B (the probability to take action U conditional on message B) in order to

$$\max_{q_B} \quad ((1 - \eta)(q_B + 2(1 - q_B)) + \eta(2q_B + 1 - q_B)).$$

This maximization problem is equivalent to

$$\max_{q_B} \quad (2 - \eta + q_B (2\eta - 1)),$$

and therefore, the best response correspondence of the receiver is

$$q_B^*(\eta) = \begin{cases} 1 & \text{if } \eta > \frac{1}{2} \\ [0,1] & \text{if } \eta = \frac{1}{2} \\ 0 & \text{if } \eta < \frac{1}{2}, \end{cases} \quad \text{, or,} \quad q_B^*(p_A, p_B) = \begin{cases} 1 & \text{if } p_A > p_B \\ [0,1] & \text{if } p_A = p_B \\ 0 & \text{if } p_A < p_B. \end{cases}$$

Next, we calculate the optimal mixed strategy (p_A^*, p_B^*) for the sender. To do so we consider three different cases:

<u>Case A:</u> Suppose that $p_A^* < p_B^*$. Then, it follows from the optimal behavior of the receiver that $q_A^*(p_A^*, p_B^*) = 1$ and $q_B^*(p_A^*, p_B^*) = 0$. Thus, the optimal strategy (p_A^*, p_B^*) must be the solution of the following maximization problem: Choose p_A and p_B in order to

$$\max_{p_A, p_B} \quad 0.5(2p_A + p_B + 1 - p_A + 2(1 - p_B)).$$

This maximization problem is equivalent to

$$\max_{p_A, p_B} \quad 0.5\left(3 + p_A - p_B\right).$$

But the solution to this problem is such that $p_A^* = 1$ and $p_B^* = 0$, and therefore, we have reached a contradiction. We conclude that there does not exist any equilibrium in which $p_A^* < p_B^*$.

Case B: Suppose that $p_A^* > p_B^*$. Then, it follows from the optimal behavior of the receiver that $q_A^*(p_A^*, p_B^*) = 0$ and $q_B^*(p_A^*, p_B^*) = 1$, and therefore, the optimal strategy (p_A^*, p_B^*) must be the solution of the following maximization problem: Choose p_A and p_B in order to

$$\max_{p_A, p_B} \quad 0.5\left(p_A + 2\left(1 - p_A\right) + 2p_B + 1 - p_B\right).$$

This maximization problem is equivalent to

$$\max_{p_A, p_B} \quad 0.5\left(3 - p_A + p_B\right).$$

But the solution to this problem is such that $p_A^* = 0$ and $p_B^* = 1$, and therefore, we have reached a contradiction. We conclude that there does not exist any equilibrium in which $p_A^* > p_B^*$.

Case C: Suppose that $p_A^* = p_B^*$. Then, it follows from the best response correspondences of the receiver that $q_A^* \in [0, 1]$ and $q_B^* \in [0, 1]$. Thus, the sender faces the problem

$$\begin{aligned} \max_{p_A, p_B} \quad & 0.5p_A\left(2q_A + 1 - q_A\right) + 0.5\left(1 - p_A\right)\left(2q_B + 1 - q_B\right) + \\ & 0.5p_B\left(q_A + 2\left(1 - q_A\right)\right) + 0.5\left(1 - p_B\right)\left(q_B + 2\left(1 - q_B\right)\right), \end{aligned}$$

a problem that is equivalent to

$$\max_{p_A, p_B} \quad 0.5\left(3 + p_A\left(q_A - q_B\right) + p_B\left(q_B - q_A\right)\right).$$

Hence, the best response correspondences for the sender are

$$p_A^*\left(q_A, q_B\right) = \begin{cases} 1 & \text{if } q_A > q_B \\ [0, 1] & \text{if } q_A = q_B \\ 0 & \text{if } q_A < q_B \end{cases} \text{ and } p_B^*\left(q_A, q_B\right) = \begin{cases} 1 & \text{if } q_A < q_B \\ [0, 1] & \text{if } q_A = q_B \\ 0 & \text{if } q_A > q_B. \end{cases}$$

From inspection we see that the set of mixed strategies $(p_A^*, p_B^*; q_A^*, q_B^*) = (p, p; q, q)$, where $p \in (0,1)$ and $q \in [0,1]$, can be sustained as equilibrium strategies. Finally, one can easily check that the corresponding beliefs are $\mu^*(\theta = A|m = A) = \eta^*(\theta = B|m = B) = \frac{1}{2}$.

Case 2: Suppose that $p_A^* = p_B^* = 0$. Observe from the best correspondence of the sender in case 1.C that $p_A^* = p_B^* = 0$ can only be sustained as an equilibrium strategy if $q_A^* = q_B^*$. Moreover, $p_A^* = p_B^* = 0$ implies that $\eta^*(\theta = B|m = B) = \frac{1}{2}$ and $q_B^*(\eta^*) \in [0,1]$. Since the sequential game we study consists of two-periods and the cardinality of the action space of both players is equal to two, any belief $\mu \in [0,1]$ is consistent. Yet, we obtain from the best response correspondence $q_A^*(\mu)$ in case 1 that $q_A^* = q_B^*$ if and only if $\mu = \frac{1}{2}$. Therefore, we conclude that the set of mixed strategies $(p_A^*, p_B^*; q_A^*, q_B^*) = (0, 0; q, q)$, where $q \in [0,1]$, together with the belief system $\mu^*(\theta = A|m = A) = \eta^*(\theta = B|m = B) = \frac{1}{2}$ constitutes a set of sequential equilibria.

Case 3: Suppose that $p_A^* = p_B^* = 1$. Observe from the best correspondence of the sender in case 1.C that $p_A^* = p_B^* = 1$ can only be sustained as an equilibrium strategy if $q_A^* = q_B^*$. Moreover, $p_A^* = p_B^* = 1$ implies that $\mu^*(\theta = A|m = A) = \frac{1}{2}$ and $q_A^*(\mu^*) \in [0,1]$. Although any belief $\eta \in [0,1]$ is consistent, we obtain from the best response correspondence $q_B^*(\eta)$ in case 1 that $q_A^* = q_B^*$ if and only if $\eta = \frac{1}{2}$. Therefore, we conclude that the set of mixed strategies $(p_A^*, p_B^*; q_A^*, q_B^*) = (1, 1; q, q)$, where $q \in [0,1]$, together with the belief system $\mu^*(\theta = A|m = A) = \eta^*(\theta = B|m = B) = \frac{1}{2}$ constitutes a set of sequential equilibria. $\square$

Instructions of the Punishment Game

Welcome

Thank you for coming. The purpose of this session is to study how people make decisions in a particular situation. If you have any questions, feel free to raise your hand and your question will be answered so everyone can hear. From now until the end of the session unauthorized communication of any nature with any other participant is prohibited. The experiment will be conducted through computers and all interactions between you will take place through them.

During the session you will play a game that gives you the opportunity to make money. What you earn depends partly on your decisions and partly on the decisions of others. At the end of the session, the amount you earned will be paid to you privately in cash.

We start with a brief instruction period. During the instruction period you will be given a description of the experiment. We are about to begin.

General Instructions

In your envelope you will find a questionnaire and an official receipt. Fill in the questionnaire and write down your name and matriculation number in the receipt. You will need both forms to receive your payment at the end of the session. Your personal data will be kept confidential and will be used for statistical purposes only.

You will play a game which is repeated for 50 rounds. Before the first round, the computer will randomly divide the participants into groups of six. This division will last for the entire session. Participants within each group will play only among themselves. The assignment process is random and anonymous so you will not know who is in your group.

Next, we will go over a brief tutorial. Please interrupt at any time if you have a question.

At the beginning of each round, you will be randomly joined with another participant from your group to form a pair. In each pair, one participant is randomly chosen to be the **Sender**, and one to be the **Receiver**. Remember that this process is random and the assignment changes every round.

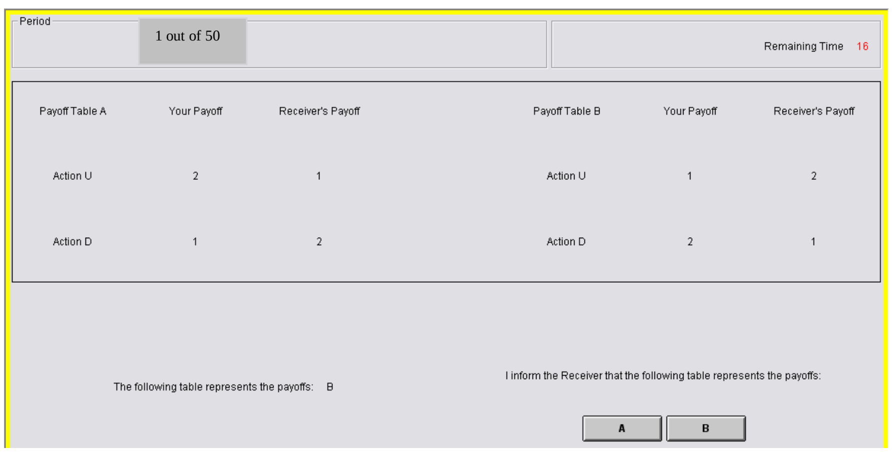
Each round, after pairs have been formed and roles have been assigned, the computer selects one of the following two payoff tables. Final payoffs for both participants will be determined according to the selected table and the action **U** or **D** taken by the Receiver later on.

Table A	Sender	Receiver
Action U	2 Points	1 Point
Action D	1 Point	2 Points

Table B	Sender	Receiver
Action U	1 Point	2 Points
Action D	2 Points	1 Point

Sender's Instructions

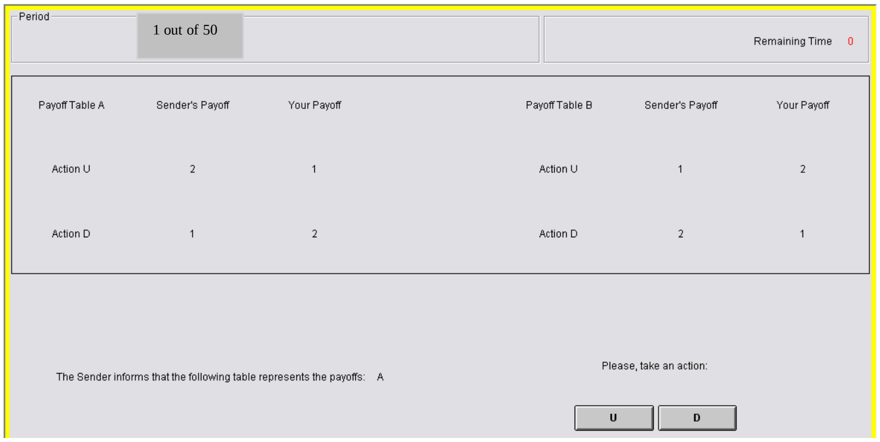
At the beginning of the round **only** the Sender will be informed about the actual payoff table chosen by the computer. The Sender is the first one to take a decision in the game. S/He must communicate to the Receiver whether the payoff table chosen by the computer is either table **A** or table **B**. **Please, take into account that the Sender is free to tell the truth or to lie**. The computer screen for the Sender is as follows:



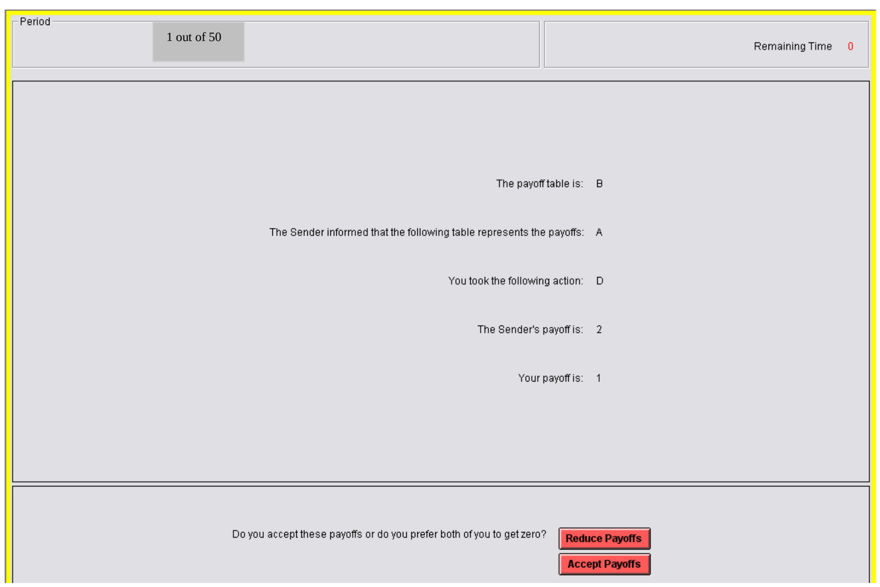
At the top of the screen you recognize the tables **A** and **B**. Below you find the information whether table **A** or table **B** was chosen by the computer (in our example it is table **B**). On the inferior right corner there are two buttons labelled **A** and **B**. By clicking on the buttons **A** or **B** you inform the Receiver that you have observed the corresponding table. The Sender has 20 seconds to take the decision. **This is the only decision the Sender takes.**

Receiver's Instructions

The Receiver takes two decisions. First, once the Receiver got the Sender's message, s/he has to decide between actions **U** and **D**. The computer screen for the Receiver is as follows:



The two tables at the top of the screen represent payoffs according to tables **A** and **B**. Below you find the message from the Sender regarding the table s/he observed (in our example the Sender has informed the Receiver that s/he observed table **A**). On the inferior right corner there are two buttons labelled **U** and **D**. By clicking on the buttons **U** or **D** you take the corresponding action. The Receiver has 20 seconds to take this decision. Once this action is taken, a new screen appears summarizing the outcome of the round so far.

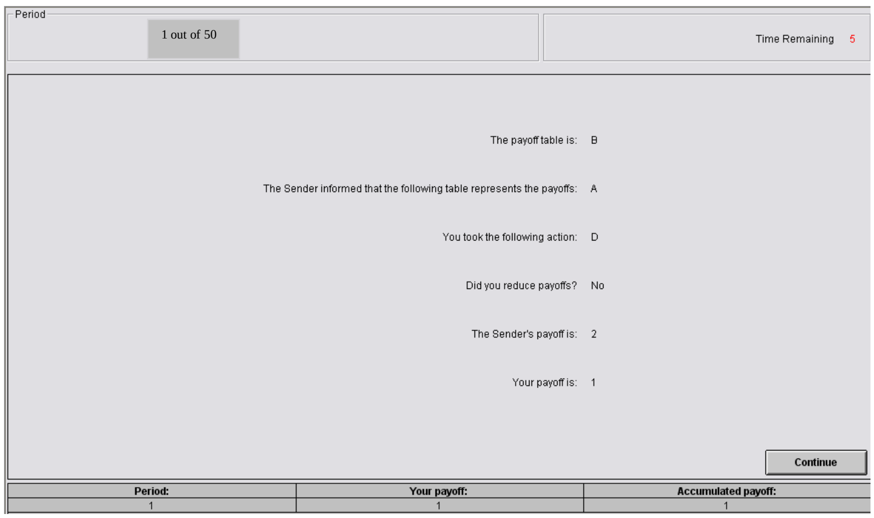


Now the Receiver is asked to take the second decision: S/He must either accept the current payoff distribution or reduce the payoff of both participants to zero. By clicking on the button **Reduce**

Payoffs or **Accept Payoffs**, the Receiver takes the corresponding action. The Receiver has 15 seconds to take this decision.

Summary of the Round

The final screen is a summary of the round: It indicates the actual payoff table, the message chosen by the Sender, the actions taken by the Receiver, and the earnings of both participants in this round. Additionally, you are also informed about your accumulated payoff.



The screen above is the Receiver's summary. It indicates that the Sender chose message **A** whereas the Receiver took action **D** and accepted the payoffs. Therefore, the Sender gets 2 Points and the Receiver 1 Point. At the end of a round, click on **Continue**. The experiment will nevertheless proceed automatically to the next round in 10 seconds.

Payment

The Points you accumulate during the course of the session will determine your payment in addition to the £**5** show-up fee. The exchange rate Points/£ is **10p per Point**. At the end of the experiment, take your questionnaire and receipt to the counter for payment. They will be matched to our computer printout. Once you are paid, you may leave.

Chapter 5

Coalition Formation in a Contest with Three Heterogenous Players

5.1 Introduction

A contest is a socio-economic environment in which players spend valuable resources in order to raise their probabilities of winning a fixed prize. Here, we analyze the incentives for cooperation of three players in the presence of the strong non-cooperative threat of a contest. In particular, we consider players who differ in their efficiency of effort or represent exogenously given groups of different size. Our main result states that a society-wide agreement may not be reached if the discrepancy between the distribution of the exogenously given group sizes and the distribution of the relative bargaining power is too high. In this case the equilibrium level of conflict is strictly positive.

Economists study contest games since the seminal work on rent-seeking by Tullock [60]. In the rent-seeking literature the individual expenditure is usually interpreted as lobbying effort in terms of time or money and the prize is taken to be a monopoly right or a license. But "lobbying" is far from being the only example, because contest games have been applied to patent races by Peréz-Castrillo and Verdier [48], to market share competition by Schmalensee [54], and to financial institutions and money by Shapley and Shubik [57]. One of the latest developments is due to Esteban and Ray [30] who concentrate on the relationship between the distribution of society into interest groups and the level of conflict defined as the social loss induced by the non-productive efforts.

Our contribution is to consider coalitions among players in a contest game. This extension is of special interest, because there is empirical evidence for the formation of coalitions in contest environments. Consider for instance a country in transition to democracy which is split into ethnic or religious groups. Often, all groups know that socially it would be best to agree on a new constitution and divide the political power, but finally, negotiation fails and a conflict between the groups emerges. A possible explanation is that if one group has a high bargaining power but is relatively small in size, then the other groups prefer to stop negotiations and fight for their political influence instead of signing an agreement in favor of the small group. The next example is an application to patent races. We observe that firms form joint ventures in research and development in order to share their knowledge and become more efficient. This means firms can raise the probability of making the next invention and to get access to a monopoly for some time through cooperation. Finally, remember the latest Soccer World Cup in Japan and South Korea which is just one example of administrations bidding jointly for the concession of a big cultural or sporting event.

Since we want to be explicit about the non-cooperative bargaining foundations of coalition formation, we use the common approach of dividing the model into two stages. In the first stage, players form coalitions and negotiate about the sharing rules of the cooperative payoffs. The coalition formation game we use is inspired by the partnership game of Gul [39] with the important difference that we allow players to exit the coalition formation game in order to free-ride on the coalition formation of the other players. In the second stage, coalitions play a contest game similar to the one proposed by Esteban and Ray [30] within the resulting coalition structure of the first stage. Finally, coalitions divide their obtained payoffs according to the sharing rules negotiated in the coalition formation game.

We solve the two stages by backward induction and determine at first the expected utility each coalition can assure itself in every possible coalition structure. Our natural prediction for this value is the unique Nash equilibrium payoff of the contest game played within the considered coalition structure. We prove in Proposition 5.1 and specially in Corollary 5.1 that the expected utility of a player who faces two single players is different from the expected utility s/he would get if the other two players have formed a coalition. Because of this externality the underlying contest game is a partition function game. In the next step, we solve the coalition formation game for the equilibrium coalition structures and the equilibrium expected utilities using stationary strategies. We show in Proposition 5.2 that if the relative efficiency of effort is distributed in the same way as the relative bargaining power, then the grand coalition forms. Moreover, if players are sufficiently patient, then every player receives in equilibrium her/his relative efficiency of effort. Finally, we show by means of a graphical analysis and an example that the grand coalition is no longer the unique equilibrium coalition structure if every player has the same bargaining power. Therefore, we may observe a strictly positive level of effort in equilibrium which to our knowledge is a new result in models of coalition formation based on bargaining.

Few papers have analyzed the question of coalition formation in contest games. Tan and Wang [59] study the formation of alliances in a rent-seeking model with heterogenous players and exogenously given effort levels. They show in a model of repeated conflict (if a coalition of more than one player wins the prize, then the members of the alliance compete further until a unique winner is determined) that an equilibrium coalition structure exists for the case of three individuals and that in this case generally the two weak individuals form a coalition against the strong one. Baik and Lee [4] and [5] study a rent-seeking model with a linear cost function. They use the open membership game as coalition formation game and obtain that coalitions with about fifty percent of the individuals are formed. Esteban and Sákovics [31] consider a model of repeated conflict with endogenous effort and bilateral coalition formation. For the case of three individuals their results predict the formation of a coalition of size two. Finally, Bloch et al. [10] study the endogenous formation of coalitions in a simplified version of the contest game of Esteban and Ray [30] with a quadratic cost function. Their main result states that the grand coalition is the unique equilibrium coalition structure of the size announcement game by Bloch [9]. Since a finite number of individuals are assumed to be identical, two individuals who belong to the same coalition receive the same share of the coalitional payoff. Therefore, the prize is divided equally among all individuals and the equilibrium level of conflict is zero.

The remainder is structured as follows: In the next Section, we introduce the contest game and derive the partition function game. In Section 5.3, we solve the bargaining model, and in Section 5.4, we discuss some of our modelling choices. The proof of Proposition 5.2 can be found in the Appendix.

5.2 Contest as a Partition Function Game

Consider three individuals who fight over a prize with a value normalized to 1. Let $N = \{1, 2, 3\}$ be the set of individuals. A coalition C is a nonempty subset of N. A coalition structure π is a partition of N. The set of all coalition structures is denoted by Π. Let $V(C, \pi)$ be the worth of coalition C in π.

We start by describing the expected utility maximization problem for the generic individual of the coalition structure $\pi = \{1|2|3\}$. Every individual i makes at the same time and independently of the others a non-productive effort $r_i \in \mathbb{R}_+$. The efficiency of effort of individual i is common knowledge and denoted by $n_i > 0$. We normalize the parameter vector $\mathbf{n} = (n_1, n_2, n_3)$ to $\sum_{i=1}^{3} n_i = 1$. We order the individuals by assuming that $n_1 \geq n_2 \geq n_3$ and define the total level of conflict as $R = \sum_{j=1}^{3} n_j r_j$. The probability that individual i wins the contest is assumed to be of the proportional form

$$p_i(\mathbf{r}; \mathbf{n}) = \frac{n_i r_i}{\sum_{j=1}^{3} n_j r_j}.$$

If $r_j = 0$ for all j, then $p_i(\mathbf{r}; \mathbf{n}) = n_i$ by convention. Moreover, let the cost of effort be equal to the level of effort.[1] Therefore, the expected utility maximization problem for

[1] We make this assumption for purely technical reasons, because if we had considered a convex cost function, then the calculation of the Nash equilibrium in the coalition structure $\pi = \{1|2|3\}$ would have become far too complicated.

individual i given r_{-i} is to choose $r_i \in \mathbb{R}_+$ in order to

$$\max_{r_i \geq 0} \left(\frac{n_i r_i}{\sum_{j=1}^{3} n_j r_j} - r_i \right). \tag{5.1}$$

This part of our model is similar to a special case of the contest model proposed by Esteban and Ray [30].[2] In their model the cost of effort is a convex function, whereas we consider the linear cost function suggested by Tullock [60]. The difference of our model to the latter one stems from the fact that Tullock's analysis is restricted to the case of identical individuals.

Consider now the coalition structure $\pi = \{ij\,|k\}$. We assume that if a coalition forms, then the members of a coalition agree to expend the same amount of effort and to choose the common effort level in order to maximize the coalitional payoff.[3] Therefore, the coalition $\{i, j\}$ has the following expected utility maximization problem: given r_k, choose $r_{ij} \in \mathbb{R}_+$ in order to

$$\max_{r_{ij} \geq 0} \left(\tfrac{(n_i+n_j)r_{ij}}{(n_i+n_j)r_{ij}+n_k r_k} - r_{ij} \right). \tag{5.2}$$

Accordingly, the expected utility maximization problem for individual k is: given r_{ij}, choose $r_k \in \mathbb{R}_+$ in order to

$$\max_{r_k \geq 0} \left(\tfrac{n_k r_k}{(n_i+n_j)r_{ij}+n_k r_k} - r_k \right). \tag{5.3}$$

If $r_k = r_{ij} = 0$, then $p_{ij}\left(r_{ij}, r_k; \mathbf{n}\right) = n_i + n_j$ and $p_k\left(r_{ij}, r_k; \mathbf{n}\right) = n_k$ by convention.

So far we have described two of the three possible types of coalition structures. The optimal decision of the grand coalition is to put zero effort, because it receives the private good anyway. Therefore, it has a worth of 1.

Suppose that the coalition structure π is the outcome of the coalition formation game. Since the contest game is a simultaneous move game, we take the Nash equilibrium of the

[2]Our interpretation of the parameter vector $\mathbf{n}$ is not the same as the one of Esteban and Ray [30], because they regard n_i as the relative size of the exogenously given group i. We discuss this point in detail in the last Section of the paper.

[3]We could define the objective function of the coalition $\{i, j\}$ by assuming that individuals do not make the same amount of effort, an argument which has been brought forward by Bloch et al. [10]. We provide evidence that our main results are invariant with respect to this change in our model later on.

contest game played within π as the natural prediction of the effort vector. Proposition 5.1 characterizes the unique Nash equilibrium for every non-trivial coalition structure.

Proposition 5.1. *(a) The unique Nash equilibrium* (r_1^*, r_2^*, r_3^*) *in the coalition structure* $\pi = \{1\,|2|\,3\}$ *is as follows: (a.1) if* $n_3 > 0.25$*, then* $r_i^* = \frac{2n_j n_k (n_i n_j + n_i n_k - n_j n_k)}{(n_i n_j + n_i n_k + n_j n_k)^2}$ *for all* i*; (a.2) if* $n_3 \leq 0.25$*, then* $r_3^* = 0$ *and* $r_1^* = r_2^* = \frac{n_1 n_2}{(n_1+n_2)^2}$*. (b) The unique Nash equilibrium* $\left(r_{ij}^*, r_k^*\right)$ *in the coalition structure* $\pi = \{ij\,|k\}$ *is* $r_{ij}^* = r_k^* = (n_i + n_j)\, n_k$*.*

Proof. a) Consider the coalition structure $\pi = \{1\,|2|\,3\}$ and in particular, the maximization problem (5.1) for player i. From the first order condition we obtain

$$\frac{n_i R - n_i^2 r_i}{R^2} - 1 = 0 \Leftrightarrow \frac{n_i}{R}\left(\frac{R}{R} - \frac{n_i r_i}{R}\right) = 1 \Leftrightarrow p_i = 1 - \frac{R}{n_i}. \tag{5.4}$$

Since the level of conflict in equilibrium is implicitly given by the equation $\sum_{i=1}^{3} p_i^* = 1$, the equation $\sum_{i=1}^{3}\left(1 - \frac{R^*}{n_i}\right) = 1$ must be satisfied. Straightforward calculus yields

$$R^* = \frac{2n_1 n_2 n_3}{n_1 n_2 + n_1 n_3 + n_2 n_3} = \frac{2}{m},$$

where $m \equiv \frac{n_1 n_2 + n_1 n_3 + n_2 n_3}{n_1 n_2 n_3}$ is the harmonic mean of $\mathbf{n}$. Plug R^* into the last equation of (5.4) to obtain $p_i^* = \frac{n_i n_j + n_i n_k - n_j n_k}{n_i n_j + n_i n_k + n_j n_k}$. Rewrite the first equality of (5.4) in terms of r_i to deduce that

$$r_i^* = \frac{n_i R^* - (R^*)^2}{n_i^2} = \frac{n_i 2 n_i n_j n_k (n_i n_j + n_i n_k + n_j n_k) - 4n_i^2 n_j^2 n_k^2}{n_i^2 (n_i n_j + n_i n_k + n_j n_k)^2} = \frac{2n_j n_k (n_i n_j + n_i n_k - n_j n_k)}{(n_i n_j + n_i n_k + n_j n_k)^2}. \tag{5.5}$$

Hence, r_i^* is positive if and only if $n_i n_j + n_i n_k - n_j n_k > 0$. We rewrite the condition for player 1 and 2 as $n_1 n_2 + n_3 (n_1 - n_2) > 0$ and $n_2 n_3 + n_1 (n_2 - n_3) > 0$, respectively. Since $n_1 \geq n_2 \geq n_3$, it is obvious that both conditions are satisfied. Finally, consider the condition for player 3 which can be stated as $n_3 > \frac{n_1 n_2}{n_1 + n_2}$. Since the weak inequality $\frac{1}{4}(t_1 + t_2) \geq t_1 t_2$ holds for any $t_1, t_2 \in [0, 1]$, we have as a particular case $\frac{1}{4}(n_1 + n_2) \geq n_2 n_1$. Hence, $n_3 > 0.25$ is a necessary condition for $\mathbf{r}^* > 0$. Evaluating the second order condition at the unique critical point yields

$$\frac{\left(n_i^2 - n_i^2\right)(R^*)^2 - 2R^* n_i \left(n_i R^* - n_i^2 r_i^*\right)}{R^4} < 0,$$

where the inequality holds because of $R^* - n_i r_i^* = \sum_{j \neq i} n_j r_j^* > 0$. Hence, if $n_3 > 0.25$, then $r_i^* = \frac{2n_j n_k (n_i n_j + n_i n_k - n_j n_k)}{(n_i n_j + n_i n_k + n_j n_k)^2}$ for all i constitutes the unique Nash equilibrium of the contest game in the coalition structure $\pi = \{1\,|2|\,3\}$.

Suppose that $n_3 \leq 0.25$. Then $r_3^* = 0$, and the first order condition of the maximization problem (5.1) for player i, given $r_3 = 0$ and r_j, is $\frac{n_i}{n_i r_i + n_j r_j}(1 - p_i) = 1$. We rewrite it as $p_i^*(1 - p_i^*) = r_i^*$. We use $p_3^* = 0$ and obtain $p_1^* p_2^* = r_1^* = r_2^*$. Finally, because $p_i^* = \frac{n_i r_i^*}{n_i r_i^* + n_j r_j^*}$, we verify that $r_1^* = r_2^* = \frac{n_1 n_2}{(n_1 + n_2)^2}$. Since the first order condition is as well sufficient we have shown that if $n_3 \leq 0.25$, then the vector $(r_1^*, r_2^*, r_3^*) = \left(\frac{n_1 n_2}{(n_1+n_2)^2}, \frac{n_1 n_2}{(n_1+n_2)^2}, 0\right)$ constitutes the unique Nash equilibrium of the contest game in the coalition structure $\pi = \{1\,|2|\,3\}$.

(b) We turn now to the coalition structure $\pi = \{ij\,|k\}$. The first order condition of the maximization problems (5.2) and (5.3) are

$$\frac{n_i + n_j}{(n_i + n_j) r_{ij} + n_k r_k}(1 - p_{ij}) - 1 = 0 \quad \text{and} \quad \frac{n_k}{(n_i + n_j) r_{ij} + n_k r_k}(1 - p_k) - 1 = 0. \tag{5.6}$$

We multiply the first equation of (5.6) by r_{ij} and the second one by r_k and obtain the condition $p_{ij}^* p_k^* = r_{ij}^* = r_k^*$. In the next step, we use the definitions of p_{ij} and p_k in order to get that $r_{ij}^* = (n_i + n_j)\, n_k$. Finally, we evaluate the second order conditions in the corresponding critical points. Since it is easy to verify that the conditions $-\frac{2(n_i+n_j)^2 n_k r_k^*}{\left((n_i+n_j) r_{ij}^* + n_k r_k^*\right)^3} < 0$ and $-\frac{(n_i+n_j) n_k^2 r_{ij}^*}{\left((n_i+n_j) r_{ij}^* + n_k r_k^*\right)^3} < 0$ hold, we have shown that the vector $\left(r_{ij}^*, r_k^*\right) = \left((n_i + n_j)\, n_k, (n_i + n_j)\, n_k\right)$ constitutes the unique Nash equilibrium in the coalition structure $\pi = \{ij\,|k\}$. □

We comment on Proposition 5.1 with respect to two points. First, the corner equilibrium in the coalition structure $\pi = \{1\,|2|\,3\}$ for $n_3 \leq 0.25$ is due to the assumption of a linear cost function (a sufficient condition assuring the existence of an interior equilibrium is $\lim_{r_i \to 0} c'(r_i) \to 0$). The intuition of this result is as follows: If the efficiency of effort of individual 3 is rather small and individual 1 and 2 play best responses, then individual 3 has a negative expected payoff by exerting a strictly positive level of effort because the costs of a small effort do not tend sufficiently fast to zero given the negligible probability of winning the prize. Hence, individual 3 prefers not to make any effort at all, and, as a consequence, the other two individuals behave as if the third individual would not exist. Second, notice that coalition $\{i,j\}$ and individual k exert the same level of effort in the unique equilibrium in the coalition structure $\pi = \{ij\,|k\}$. This is a known result for the rent seeking model of Tullock [60] with two identical individuals. The difference is that due to the heterogeneity factor, coalition $\{i,j\}$ and individual k do not have any more the same winning probability which implies that the expected utilities differ.

We derive the partition function game V from Proposition 5.1 by plugging the equilibrium efforts for every type of coalition structure into the corresponding objective functions.

Corollary 5.1. *The partition function game V is equal to*

$$\begin{aligned}
V(123,\{123\}) &= 1\\
V(ij,\{ij\,|k\}) &= (n_i+n_j)^2\\
V(k,\{ij\,|k\}) &= n_k^2\\
V(k,\{i\,|j|\,k\}) &= \begin{cases} \left(1-\frac{2n_in_j}{n_in_j+n_in_k+n_jn_k}\right)^2 & \text{if } n_3 > 0.25\\ \left(\frac{n_k}{n_1+n_2}\right)^2 & \text{if } n_3 \le 0.25 \text{ and } k \ne 3\\ 0 & \text{if } n_3 \le 0.25 \text{ and } k = 3. \end{cases}
\end{aligned}$$

5.3 The Coalition Formation Game

Since V summarizes all necessary information of the contest game, we are ready to address the question of coalition formation. As it has already been outlined, the main difference between our coalition formation game and the partnership game of Gul [39] is that our game allows coalitions to opt out, a change which is crucial for deriving non-efficient outcomes. The game is parameterized by a common discount factor $0 < \delta < 1$ and a probability vector $\mathbf{q} = (q_1, q_2, q_3)$ representing the relative bargaining power of the players.

The Bilateral Bargaining Game

Period 0:

Players decide sequentially and publicly whether to stay in the coalition formation game or to leave it according to the ordering $1, 2, 3$. Every player who exits the coalition formation game forms a coalition on his own. Let S_0 be the set of players who decide to stay and denote by s_0 the cardinality of S_0. If $s_0 \leq 1$, then the contest game is played within the coalition structure $\{1\,|2|\,3\}$. If $s_0 \geq 2$, then a randomly selected bilateral meeting among the players in S_0 takes place. We assume that every possible meeting occurs with equal probability. Suppose that i and j meet each other. Player i is chosen with probability $q_{i,j} = \frac{q_i}{q_i+q_j}$ to make an offer $x^0_{i,j} \in \mathbb{R}_+$ which can be accepted or rejected by j. The offer describes j's share of the payoff $V(ij, \{ij\,|k\})$. If j rejects $x^0_{i,j}$, then we set $\pi^0 = \{1\,|2|\,3\}$ and pass to the next period. If j accepts $x^0_{i,j}$, then the coalition $\{i, j\}$ forms and the actual coalition structure becomes $\pi^0 = \{ij\,|k\}$. If $k \notin S_0$, then the process of coalition formation stops and the contest game is played within the coalition structure π^0. If $k \in S_0$, then we pass to the next period.

Period t:

The game arrives at period $t > 0$ if (a) $\pi^{t-1} = \{1\,|2|\,3\}$ and $s_{t-1} \geq 2$, or if (b) $\pi^{t-1} = \{ij\,|k\}$ and $s_{t-1} = 3$. Players in S_{t-1} decide sequentially and publicly according to the ordering $1, 2, 3$ restricted to S_{t-1} whether to stay or to leave the coalition formation game. Let S_t be the set of players who decide to stay and denote by s_t the cardinality of S_t.

Suppose that $\pi^{t-1} = \{1\,|2|\,3\}$. If $s_t \leq 1$, then the contest game is played within the coalition structure π^{t-1}. If $s_t \geq 2$, then a randomly selected bilateral meeting among players in S_t takes place. Every possible meeting occurs with equal probability. Suppose that i and j meet each other. Player i is chosen with probability $q_{i,j} = \frac{q_i}{q_i+q_j}$ to make an offer of $x^t_{i,j} \in \mathbb{R}_+$ which can be accepted or rejected by j. If j rejects $x^t_{i,j}$, then we set $\pi^t = \pi^{t-1}$ and pass to the next period. If j accepts $x^t_{i,j}$, then the coalition $\{i, j\}$ forms and the actual coalition structure becomes $\pi^t = \{ij\,|k\}$. If $k \notin S_t$, then the process of coalition formation stops and the contest game is played within the coalition structure π^t. If $k \in S_t$, then we pass to the next period.

Suppose that $\pi^{t-1} = \{ij\,|k\}$. If $s_t < 3$, then the final coalition structure is π^{t-1} and the corresponding contest game is played. If $s_t = 3$, then player i, who represents the coalition $\{i, j\}$ and is the first in the ordering between i and j generated by 1,2,3, and player k meet. Player i is chosen with probability $q_{ij,k} = q_i + q_j$ to make an offer $x^t_{ij,k} \in \mathbb{R}_+$ which can be accepted or rejected by k. The offer describes k's share of the payoff $V(123, \{123\})$. If k rejects $x^t_{ij,k}$, then we set $\pi^t = \pi^{t-1}$ and pass to the next period. If k accepts $x^t_{ij,k}$, then the grand coalition forms and payoffs are assigned accordingly.

Figure 5.1 below represents the Bilateral Bargaining game for $t = 0$ and $t = 1$.

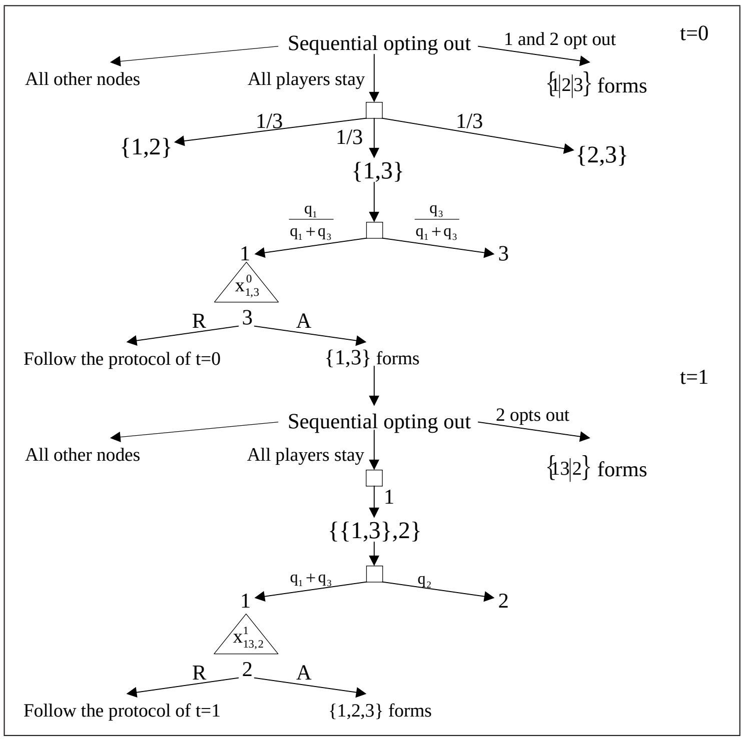


Figure 5.1: The Bilateral Bargaining Game

We would like to characterize the equilibrium coalition structures and utilities for any arbitrary vector $\mathbf{q}$. Unfortunately, our parameter space would be enlarged too much, because our results have to rely partly on a graphical analysis even for a fixed $\mathbf{q}$. Therefore, we analyze the bilateral bargaining game for the probability vectors (n_1, n_2, n_3) and $\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$. The first probability vector is focal, because it reflects symmetry between bargaining power and efficiency of effort, the other one has been studied by Gul [39].

Proposition 5.2. *Suppose that* $\mathbf{q} = \mathbf{n}$*. For all* $\delta < 1$*, the bilateral bargaining game has a unique stationary subgame perfect equilibrium outcome. Let* $U_i^*(\delta)$ *be the expected utility of player* i *in the equilibrium corresponding to* δ*. Then* $\lim_{\delta \to 1} U_i^*(\delta) = n_i$*. Moreover, the grand coalition is the unique equilibrium coalition structure.*

The detailed proof can be found in the Appendix and is sketched now: We solve the coalition formation game by backward induction. To do so suppose that the offer $x_{i,j}^t$ has been accepted by j in period t and that coalition $\{i, j\}$ and individual k are still in the coalition formation game at $t+1$. Then, it is easy to show that the offers $x_{k,ij}^{t+1}$ and $x_{ij,k}^{t+1}$ are acceptable offers (Lemma 5.2) and that is optimal for coalition $\{i, j\}$ and individual k to make offers which leaves the other party indifferent between accepting and rejecting it. Given the optimal offers $x_{ij,k}^* = \delta^2 n_k$ and $x_{k,ij}^* = \delta^2 (n_i + n_j)$ it is possible to calculate the one-period discounted expected utilities ($U_i^*(\delta, x_{i,j}^t) + U_i^*(\delta, x_{i,j}^t) = \delta(n_i + n_j)$ and $U_k^*(\delta, x_{i,j}^t) = \delta n_k$) and to see that no individual has incentives to drop out of the coalition formation game in the beginning of period $t+1$ (Lemma 5.3). Therefore, the grand coalition is the unique coalition structure to be observed after the formation of coalition $\{i, j\}$ when individual k has not opted out before. Next, we solve for the optimal strategies at t and analyze whether coalition $\{i, j\}$ really forms. If no individual has left the coalition formation game until the beginning of period t, then Nature chooses among six different options, because there are three different coalitions with two different proposers each. Again it is optimal for i to make an offer $x_{i,j}^t$ which leaves j indifferent between accepting and rejecting it, an observation which allows us to derive a linear system of three equations with the expected utilities as unknown variables (Lemma 5.4). At this stage every individual incorporates the knowledge that the grand coalition forms for sure afterwards. The solution of the linear system is such that every individual gets his efficiency of effort as the discount rate tends to 1. Finally, we check that no individual wants to opt out in the beginning of period t (Lemma 5.5).

Proposition 5.2 is an efficiency result which provides a planner with important information how to minimize conflict in society. But since there is empirical evidence for a positive level of conflict, we show next that it is possible to sustain other coalition structures in equilibrium if the vector of relative bargaining power $\mathbf{q}$ is sufficiently different from the vector of relative efficiency $\mathbf{n}$. With respect to the examples stated before we have the following situations in mind:

- There are countries where big ethnic groups have little political power. In certain situations, these groups are not willed to negotiate over political decisions with the ruling party and rather decide to fight for their preferred outcome.

- New biotechnological firms entering a market can be very efficient in developing patents and medicaments. Often, these firms form joint ventures with incumbents that have a high bargaining power due to their financial assets.

- Japan and South Korea have little influence within the FIFA (Fédération Internationale de Football Association) who decides where to organize the Soccer World Cup. On the other hand, both countries are able to make big marketing efforts and provide up to date infrastructures.

In order to derive the inefficiency result we consider now the case when $\mathbf{q} = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$. Since the calculations for characterizing the stationary subgame perfect equilibrium of the corresponding bilateral bargaining game become even much longer, we do not present a formal derivation of the result. Rather, we concentrate on a graphical representation and calculate a detailed example.

In Figure 5.2, the set of points (n_1, n_2, n_3) fulfilling the constraints $n_1 \geq n_2 \geq n_3$ and $n_1 + n_2 + n_3 = 1$ are the ones lying within the triangle indicated by the thicker lines. The grand coalition forms for sure for all combinations of points (n_1, n_2, n_3) lying in non-shaded area of the triangle. The lightly grey shaded area within the triangle corresponds to the set of points (n_1, n_2, n_3) for which the grand coalition forms with probability $\frac{2}{3}$. The grand coalition forms with probability $\frac{1}{3}$ for the set of points (n_1, n_2, n_3) lying in the darkly grey shaded area within the triangle. The black shaded area within the triangle corresponds to set of points (n_1, n_2, n_3) for which the grand coalition does not form.

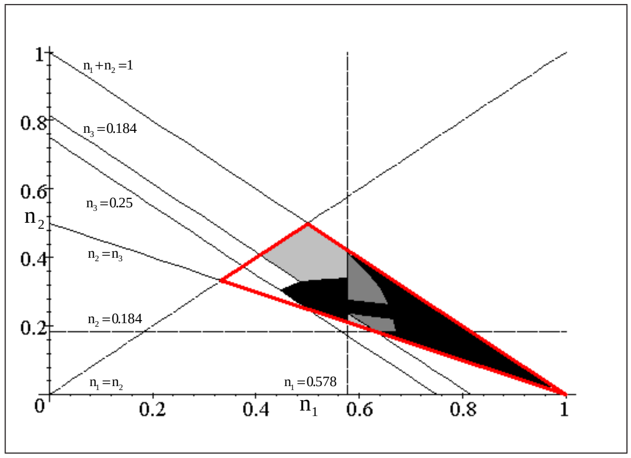


Figure 5.2: Equilibrium Coalition Structures

Example 5.1. Let $\mathbf{n} = (\frac{8}{10}, \frac{1}{10}, \frac{1}{20})$. The partition function game V is equal to

$$\begin{array}{lll} V(123, \{123\}) = 1 & & \\ V(12, \{12\,|3\}) = 0.81 & V(13, \{13\,|2\}) = 0.81 & V(23, \{23\,|1\}) = 0.04 \\ V(3, \{12\,|3\}) = 0.01 & V(2, \{13\,|2\}) = 0.01 & V(1, \{23\,|1\}) = 0.64 \\ V(1, \{1\,|2|\,3\}) = 0.79 & V(2, \{1\,|2|\,3\}) = 0.015 & V(3, \{1\,|2|\,3\}) = 0. \end{array}$$

Suppose that $\delta \to 1$. We solve the coalition formation game using stationary strategies for the vector of relative bargaining powers $\mathbf{q} = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$. If j has accepted the offer $x^t_{i,j}$ and if $s_{t+1} = 3$, then the optimal offers in period $t+1$ are accepted and are such that $x^*_{ij,k} = \frac{1}{3}$ and $x^*_{k,ij} = \frac{2}{3}$. Therefore, in this subgame the one-period discounted utilities are equal to

$$\begin{aligned} U^*_i(x^t_{i,j}) = &\ \tfrac{2}{3}(1 - x^*_{ij,k})(1 - x^t_{i,j}) + \tfrac{1}{3}x^*_{k,ij}(1 - x^t_{i,j}) = \tfrac{2}{3}(1 - x^t_{i,j}) \\ U^*_j(x^t_{i,j}) = &\ \tfrac{2}{3}(1 - x^*_{ij,k})x^t_{i,j} + \tfrac{1}{3}x^*_{k,ij}x^t_{i,j} = \tfrac{2}{3}x^t_{i,j} \\ U^*_k(x^t_{i,j}) = &\ \tfrac{2}{3}x^*_{ij,k} + \tfrac{1}{3}(1 - x^*_{k,ij}) = \tfrac{1}{3}. \end{aligned}$$

We check that (a) if $\{i,j\} = \{1,2\}$ or if $\{i,j\} = \{1,3\}$, then coalition $\{i,j\}$ leaves the coalition formation game after its formation and the coalition structure $\pi = \{ij\,|k\}$ forms, and (b) if $\{i,j\} = \{2,3\}$, then individual 1 opts out of the coalition formation game and the final coalition structure is equal to $\pi = \{23\,|1\}$. This information is incorporated by the individuals in order to determine their optimal offer in period t. If $s_t = 3$, then i makes an offer $x_{i,j}^*$ which is such that $x_{i,j}^* = U_j^*$, where U_j^* is the expected utility of the game being equal to the expected continuation utility from rejecting the offer. In particular,

$$\begin{aligned} U_1^* &= \tfrac{1}{6}\left(0.81 - U_2^*\right) + \tfrac{1}{6}\left(0.81 - U_3^*\right) + \tfrac{1}{3}0.64 + \tfrac{1}{3}U_1^* \\ U_2^* &= \tfrac{1}{6}\left(0.81 - U_1^*\right) + \tfrac{1}{6}\left(0.04 - U_3^*\right) + \tfrac{1}{3}0.01 + \tfrac{1}{3}U_2^* \\ U_3^* &= \tfrac{1}{6}\left(0.81 - U_1^*\right) + \tfrac{1}{6}\left(0.04 - U_2^*\right) + \tfrac{1}{3}0.01 + \tfrac{1}{3}U_3^*. \end{aligned}$$

In order to see this consider the expected utility U_1^* of individual 1. With probability $\frac{1}{6}$ individual 1 is the proposer of coalition $\{1, j\}$, $j = 2, 3$. Since coalition $\{1, j\}$ leaves the coalition formation game after its formation, the coalition gets a final payoff of 0.81 from which individual 1 has to give U_j^* to j. Nature proposes the coalition $\{2, 3\}$ with probability $\frac{1}{3}$. In this case, individual 1 opts out after the formation of the coalition and gets his stand alone value in the coalition structure $\{23\,|1\}$ which is equal to 0.64. Finally, with probability $\frac{1}{3}$ individual 1 acts as a responder and gets his expected continuation value U_1^*. The solution of the linear system of three equation is $U_1^* = 0.70$ and $U_2^* = U_3^* = 0.03$. Notice that if k is the only individual leaving the coalition formation game at t, then coalition $\{i, j\}$ splits the value $V\left(ij, \{ij\,|k\}\right)$ equally in the unique stationary subgame perfect equilibrium of the continuation game. Therefore, individuals opt out according to the following graph in the beginning of period t.

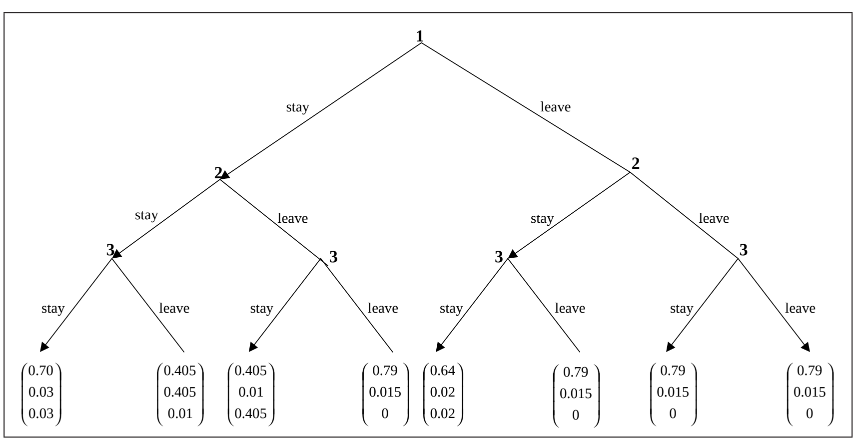


Figure 5.3: Stage t Opting Out Game

As indicated by the arrows no individual leaves the coalition in period t, and therefore, in equilibrium every two person coalition forms with probability $\frac{1}{3}$.

Finally, we comment which coalition structure forms in equilibrium in the other black shaded area in Figure 5.2. If $\mathbf{n} = (0.65, 0.25, 0.1)$ (this vector belongs to the black area divided by the dark grey areas), $\mathbf{n} = (0.5, 0.3, 0.52)$ (this is the dark area to the left) and $\mathbf{n} = (0.6, 0.35, 0.05)$ (this is the area to the upper left), then the unique equilibrium coalition structure is the one consisting of singletons.

5.4 Discussion

It is difficult to generalize our findings to more than three players, because we cannot write the worth of a coalition in a coalition structure as a function of the size of coalitions. In the literature this special function is termed "valuation". To derive it one has to concentrate on the case of identical individuals as it has been done by Bloch et al. [10]. This is the main reason why we restrict our analysis to the case of three individuals.

The next point regards the interpretation of the parameter vector $\mathbf{n}$. So far we have considered a model with three individuals who differ in the efficiency of effort. Esteban and Ray [30] assume that the prize is an excludable public good and define the parameter n_i as the relative size of the exogenously given group i. Following their assumption that all individuals who belong to the same group i are enforced by a binding agreement to make the same level of effort r_i, we interpret equation (5.1) as the expected utility maximization problem of the representative individual of group i. Accordingly, if two groups i and j form a coalition, then the relative group size of coalition $\{i, j\}$ becomes $n_i + n_j$. Hence, equation (5.2) states the expected utility maximization problem of the representative individual of coalition $\{i, j\}$.

Finally, we want to introduce a different objective function for coalition $\{i, j\}$. Bloch et al. [10] analyze a model with a finite number of homogeneous individuals. In their model individual i and j do not necessarily make the same amount of effort after the formation of coalition $\{i, j\}$ has been formed. Nonetheless, it is assumed that it is still in the interest of the individuals to maximize the joint profits. The new expected utility maximization problem of coalition $\{i,j\}$ is to take r_k as given and to choose r_i and r_j in order to

$$\max_{r_i,r_j\geq 0}\left(\frac{n_ir_i+n_jr_j}{n_ir_i+n_jr_j+n_kr_k}-r_i-r_j\right).$$

Similarly, the new expected utility maximization problem for individual k is to take r_i and r_j as given and to choose r_k in order to

$$\max_{r_k\geq 0}\left(\frac{n_kr_k}{n_ir_i+n_jr_j+n_kr_k}-r_k\right).$$

It is easy to see that if $n_i > n_j$, then $r_j^* = 0$. This result can be interpreted as a buy-out of individual j by individual i. By making all the necessary calculations we establish that the corresponding values of the partition function game become

$$V\left(ij,\{ij\,|k\}\right)=\left(\frac{\max\{n_i;n_j\}}{\max\{n_i;n_j\}+n_k}\right)^2 \quad \text{and} \quad V\left(k,\{ij\,|k\}\right)=\left(\frac{n_k}{\max\{n_i,n_j\}+n_k}\right)^2.$$

We do not show formally that Proposition 5.2 does not change due to the new values. The key point is to check whether a player wants to opt out of the coalition formation game after the formation of coalition $\{i,j\}$. All players stay in the game if and only if $n_i+n_j \geq \left(\frac{\max\{n_i;n_j\}}{\max\{n_i;n_j\}+n_k}\right)^2$ and $n_k \geq \left(\frac{n_k}{\max\{n_i;n_j\}+n_k}\right)^2$. We rewrite the first weak inequality as $\left(1-n_k\right)\left(1-n_j\right)^2 \geq \left(1-n_j-n_k\right)^2$ and reduce it to $n_i+n_j \geq n_j^2$. Since $n_j \geq n_j^2$, the weak inequality holds. We rewrite the second inequality as $n_i^2 \geq n_k\left(1-n_k-2n_i\right) = n_k\left(n_j-n_i\right)$. Since $n_i \geq n_j$ by assumption, the weak inequality holds.

Appendix

We prove Proposition 5.2 in a series of Lemmata. Since we restrict ourselves to stationary strategies, let $x_{S,T}$ be the offer made by S to T at any t. Furthermore, the partition function game V is said to be *strictly superadditive* if for all $\pi \in \Pi$ and for all $S,T \in \pi$ we have that $V\left(S\cup T,\{(\pi\backslash T\backslash S)\cup(S\cup T)\}\right) > V\left(S,\{\pi\}\right)+V\left(T,\{\pi\}\right)$.

Lemma 5.1. *The partition function game V is strictly superadditive.*

Proof. We prove that $\tilde{V}(\mathbf{n}) \equiv V(ij, \{ij|k\}) - V(i, \{i|j|k\}) - V(j, \{i|j|k\}) > 0$ for all $\mathbf{n}$ by means of a geometric argument. Suppose that $n_3 > 0.25$. In Figure 5.4 we draw the level curve $\tilde{V}(\mathbf{n}) = 0$ when $\{i, j\} = \{1, 2\}$. The three straight lines correspond to the set of points satisfying the conditions $n_1 = 0.25$, $n_2 = 0.25$ and $n_1 + n_2 = 0.75$. Therefore, the shaded triangle in the center of the figure is the set of points (n_1, n_2, n_3) where $n_1 > 0.25$, $n_2 > 0.25$ and $n_1 + n_2 < 0.75$. Notice that this area is bounded away from the level curve at zero. Since $\tilde{V}(\mathbf{n})$ is a continuous function in $\mathbf{n}$, the result follows if we find a point (n_1, n_2, n_3) in the shaded area for which $\tilde{V}(\mathbf{n}) > 0$. If $n_i = \frac{1}{3}$ for all i, then $\tilde{V}\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right) = \frac{4}{9} - \frac{2}{9} = \frac{2}{9}$ which proves the result.

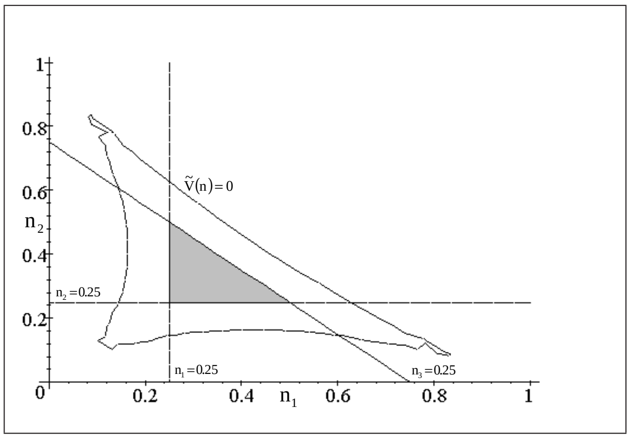


Figure 5.4: $\tilde{V}(\mathbf{n}) = 0$ when $n_3 > 0.25$

Suppose now that $n_3 \leq 0.25$. By Corollary 5.1, taking $k = 3$, we have to prove that the inequality $\tilde{V}(\mathbf{n}) = (n_1 + n_2)^2 - \frac{n_1^2 + n_2^2}{(n_1+n_2)^2} > 0$ holds. We use Figure 5.5 to establish the result. The four straight lines correspond to the conditions $n_1 + n_2 = 0.75$, $n_1 + n_2 = 1$, $n_1 = n_2$ and $n_2 = n_3$. The shaded area indicates the set of points (n_1, n_2, n_3) satisfying the conditions $0.75 < n_1 + n_2 \leq 1$ and $n_1 \geq n_2 \geq n_3$. Using the point $(n_1, n_2, n_3) = \left(\frac{2}{5}, \frac{2}{5}, \frac{1}{5}\right)$ we establish that $\tilde{V}\left(\frac{2}{5}, \frac{2}{5}, \frac{1}{5}\right) = 0.409 - 0.32 > 0$. Hence, $\tilde{V}(\mathbf{n})$ takes strictly positive values in this area whenever $n_l > 0$ for all l.

Similarly, by Corollary 5.1 and taking $i = 3$, we have to prove that $(n_3 + n_j)(1 - n_3) > n_j$. This inequality is equivalent to $n_3(1 - n_3 - n_j) = n_3 n_k > 0$. Hence, individual 3 and j profit from forming a coalition. Finally, we have to check that the inequality

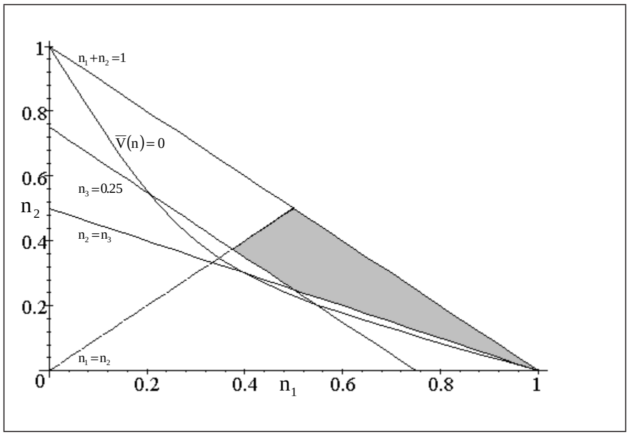


Figure 5.5: $\tilde{V}(\mathbf{n}) = 0$ when $n_3 \leq 0.25$

$V(123, \{123\}) > V(ij, \{ij\,|k\}) + V(k, \{ij\,|k\})$ holds for all possible permutations of the set of players. This follows, because the equilibrium level of conflict R^* is strictly positive in any coalition structure of the type $\pi = \{ij\,|k\}$. □

Lemma 5.2. *Let $\mathbf{q} = \mathbf{n}$ and suppose that $x_{i,j}$ has been accepted at t. If $s_{t+1} = 3$, then $x^*_{k,ij}$ and $x^*_{ij,k}$ are acceptable offers.*

Proof. Suppose that $x^*_{k,ij}$ is not an acceptable offer. If $x^*_{ij,k}$ is not an acceptable offer either, then coalition $\{i, j\}$ and player k will negotiate for ever, because by stationarity $s_{t+m} = 3$ for all $m \geq 2$. In this case the final utility of every player is zero. But if player k had left the game before, then s/he would have received an utility of $n_k^2 > 0$ from playing the contest game within the coalition structure $\pi = \{ij\,|k\}$. Hence, we have reached a contradiction to $s_{t+1} = 3$ and conclude that $x^*_{ij,k}$ must be an acceptable offer. Since $x^*_{k,ij}$ is not acceptable by assumption, the one-period discounted expected utility of player k is equal to

$$\begin{aligned} U_k\left(\delta, x^*_{i,j}\right) &= \delta\left(\left(n_i+n_j\right)x^*_{ij,k} + \delta n_k\left(n_i+n_j\right)x^*_{ij,k} + \delta^2 n_k^2\left(n_i+n_j\right)x^*_{ij,k} + \ldots\right) \\ &= \delta\left(n_i+n_j\right)x^*_{ij,k}\sum_{\tau=0}^{\infty}\delta^\tau n_k^\tau = \delta\frac{\left(n_i+n_j\right)x^*_{ij,k}}{1-\delta n_k}. \end{aligned}$$

Player k accepts the offer $x_{ij,k}$ if and only if it is at least as high as the discounted value of the expected continuation utility from rejecting it; that is, $x^*_{ij,k} \geq \delta U_k\left(\delta, x_{i,j}\right)$. On the other hand player i will not offer more than $U_k(\delta, x_{i,j})$. Thus, $x^*_{ij,k} = \frac{\delta(n_i+n_j)}{1-\delta n_k} x^*_{ij,k}$. Since $\frac{\delta(n_i+n_j)}{1-\delta n_k} \neq 1$ for all $\delta < 1$, we must have $x^*_{ij,k} = 0$. This is a contradiction to $s_{t+1} = 3$, because player k can get strictly more by leaving the game and playing the contest game within the coalition structure $\pi = \{ij\,|k\}$. □

Lemma 5.3. *Let* $\mathbf{q} = \mathbf{n}$ *and suppose that* $x_{i,j}$ *has been accepted at* t. *If* $s_t = 3$, *then the grand coalition forms and the one-period discounted expected utilities are equal to*

$$\left(U^*_i(\delta, x_{i,j}), U^*_j(\delta, x_{i,j}), U^*_k(\delta, x_{i,j})\right) = \left(\delta(n_i+n_j)(1-x_{i,j}), \delta(n_i+n_j)x_{i,j}, \delta n_k\right).$$

Proof. Suppose that $s_{t+1} = 3$. Since we know from Lemma 5.2 that $x^*_{k,ij}$ and $x^*_{ij,k}$ are acceptable offers, the one-period discounted expected utilities are given by

$$\begin{aligned}
U^*_i(\delta, x_{i,j}) &= \delta\left((n_i+n_j)(1-x_{i,j})\left(1-x^*_{ij,k}\right) + n_k(1-x_{i,j})x^*_{k,ij}\right)\\
U^*_j(\delta, x_{i,j}) &= \delta\left((n_i+n_j)x_{i,j}\left(1-x^*_{ij,k}\right) + n_k x_{i,j} x^*_{k,ij}\right)\\
U^*_k(\delta, x_{i,j}) &= \delta\left((n_i+n_j)x^*_{ij,k} + n_k\left(1-x^*_{k,ij}\right)\right).
\end{aligned} \tag{5.7}$$

Player i accepts $x^*_{k,ij}$ if and only if $(1-x_{i,j})x^*_{k,ij} \geq \delta U^*_i(\delta, x_{i,j})$. Therefore, in equilibrium the equation must be satisfied with equality. Using a similar argument we establish that $x^*_{ij,k} = \delta U^*_k(\delta, x_{i,j})$. The solution of the system of linear equations (5.7), given $x_{i,j}$, is

$$\begin{aligned}
&\left(x^*_{ij,k}, x^*_{k,ij}\right) = \left(\delta^2 n_k, \delta^2(n_i+n_j)\right)\\
&\left(U^*_i(\delta, x_{i,j}), U^*_j(\delta, x_{i,j}), U^*_k(\delta, x_{i,j})\right) = \left(\delta(n_i+n_j)(1-x_{i,j}), \delta(n_i+n_j)x_{i,j}, \delta n_k\right).
\end{aligned}$$

This would be an equilibrium if it is optimal for every player to stay in the game after the formation of coalition $\{i,j\}$. If player k had left the game before, then s/he would have received an expected utility of δn_k^2 which is strictly less than δn_k. If player j had opted out, then s/he would have received an expected utility of $\delta(n_i+n_j)^2 x_{i,j}$ which is strictly less than $U^*_j(\delta, x_{i,j})$. Finally, player i does not to opt out either, because if s/he did so, then her/his expected utility would be equal to $\delta(n_1+n_2)^2(1-x_{i,j})$. But this is strictly less than $U^*_i(\delta, x_{i,j})$. □

Lemma 5.4 *Let* $\mathbf{q} = \mathbf{n}$ *and suppose that* $s_t = 3$. *Then the one-period discounted expected utility of player* l *is equal to* $U^*_l(\delta) = \delta n_l$ *for all* $l = 1, 2, 3$.

Proof. If player j accepts the offer $x_{i,j}$ at t, then $U_i^*(\delta, x_{i,j}) + U_j^*(\delta, x_{i,j}) = \delta(n_i + n_j)$. The corresponding stand alone expected utility of individual k is $U_k^*(\delta, x_{i,j}) = \delta n_k$. Since the final utility of coalition $\{i, j\}$ is independent of the applied sharing rule, player i selects the offer that makes individual j indifferent between accepting and rejecting it. That is, $\delta(n_i + n_j) x_{i,j}^* = \delta U_j^*(\delta)$, where $U_j^*(\delta)$ is the expected utility of player j at the beginning of stage t. Hence, the share which remains for player i is equal to $\delta(n_i + n_j)\left(1 - x_{i,j}^*\right) = \delta(n_i + n_j) - \delta U_j^*(\delta)$. Player i meets player j and is chosen to make the offer with probability $\frac{1}{3}\frac{n_i}{n_i+n_j}$. Player j and k meet with probability $\frac{1}{3}$. In this case player i gets his stand alone value δn_i. Finally, player i meets j in the role of the responder with probability $\frac{1}{3}\frac{n_j}{n_i+n_j}$. Therefore, the expected utilities of the players are

$$
\begin{aligned}
U_1^*(\delta) = {} & \tfrac{1}{3}\tfrac{n_1}{n_1+n_2}\left[\delta(n_1+n_2) - \delta U_2^*(\delta)\right] + \tfrac{1}{3}\tfrac{n_1}{n_1+n_3}\left[\delta(n_1+n_3) - \delta U_3^*(\delta)\right] + \tfrac{1}{3}\delta n_1 + \\
& + \tfrac{1}{3}\left(\tfrac{n_2}{n_1+n_2} + \tfrac{n_3}{n_1+n_3}\right)\delta U_1^*(\delta) \\
U_2^*(\delta) = {} & \tfrac{1}{3}\tfrac{n_2}{n_1+n_2}\left[\delta(n_1+n_2) - \delta U_1^*(\delta)\right] + \tfrac{1}{3}\tfrac{n_2}{n_2+n_3}\left[\delta(n_2+n_3) - \delta U_3^*(\delta)\right] + \tfrac{1}{3}\delta n_2 + \\
& + \tfrac{1}{3}\left(\tfrac{n_1}{n_1+n_2} + \tfrac{n_3}{n_2+n_3}\right)\delta U_2^*(\delta) \\
U_3^*(\delta) = {} & \tfrac{1}{3}\tfrac{n_3}{n_1+n_3}\left[\delta(n_1+n_3) - \delta U_1^*(\delta)\right] + \tfrac{1}{3}\tfrac{n_3}{n_2+n_3}\left[\delta(n_2+n_3) - \delta U_2^*(\delta)\right] + \tfrac{1}{3}\delta n_3 + \\
& + \tfrac{1}{3}\left(\tfrac{n_1}{n_1+n_3} + \tfrac{n_2}{n_2+n_3}\right)\delta U_3^*(\delta).
\end{aligned}
$$

The solution of the system of three linear equation and three unknowns is $U_l^*(\delta) = \delta n_l$ for all $l = 1, 2, 3$. □

Lemma 5.5. *Let $\mathbf{q} = \mathbf{n}$. Then, no player leaves the game at t.*

Proof. Assume that only player k leaves the game at t. Since the partition function game V is strictly superadditive by Lemma 5.1, we can apply similar arguments to the ones used in Lemmata 5.2 and 5.3 to show that player i and j form a coalition and adopt the sharing rules $\left(x_{i,j}^*, x_{j,i}^*\right) = \left(\delta^2\frac{n_j}{n_i+n_j}, \delta^2\frac{n_i}{n_i+n_j}\right)$ in the unique stationary subgame perfect equilibrium of the continuation game. This implies that the expected utilities of the players in this subgame are equal to $\left(U_i^*(\delta), U_j^*(\delta), U_k^*(\delta)\right) = \left(\delta n_i(n_i+n_j), \delta n_j(n_i+n_j), \delta n_k^2\right)$. Using all payoffs of the stationary equilibrium of the continuation games that we have already obtained, we represent in Figure 5.6 the game tree at t.

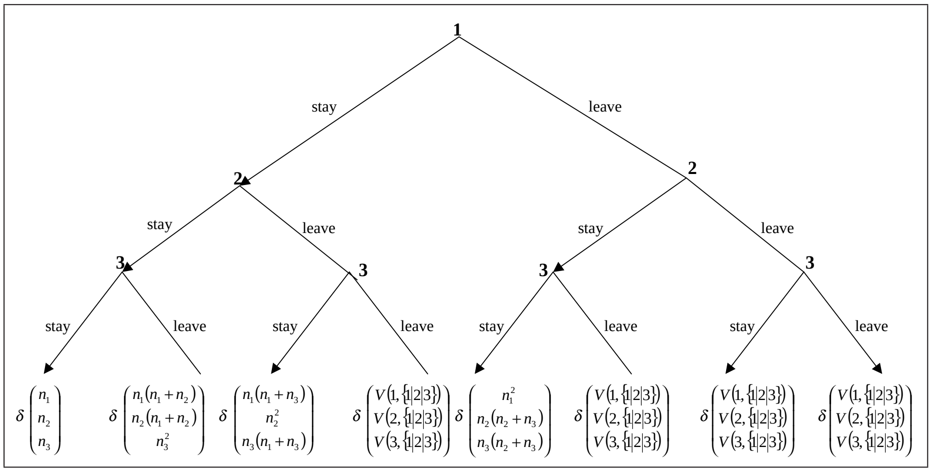


Figure 5.6: The Opting Out Game at t

We prove in the next step that player 3 may only leave the game whenever player 1 and 2 have left the game before. If $n_3 \leq 0.25$, then player 3 can do stay, because in this case $V(3, \{1|2|3\}) = 0$. Assume now that $n_3 > 0.25$. In Figure 5.7, we draw the level curve $\hat{V}(\mathbf{n}) \equiv n_i(n_i + n_j) - \left(1 - \frac{2n_j n_k}{n_i n_j + n_i n_k + n_j n_k}\right)^2 = 0$. The shaded area corresponds to the set of points (n_1, n_2, n_3) where $n_i \geq 0.25$. We check that $\hat{V}\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right) = \frac{2}{9} - \frac{1}{9} = \frac{1}{9} > 0$. Hence, by continuity of $\hat{V}(\mathbf{n})$ and since the shaded area and the indifference curve do not intersect, we have that $\hat{V}(\mathbf{n})$ takes positives values all over the area of interest. We conclude that player 3 stays in the game.

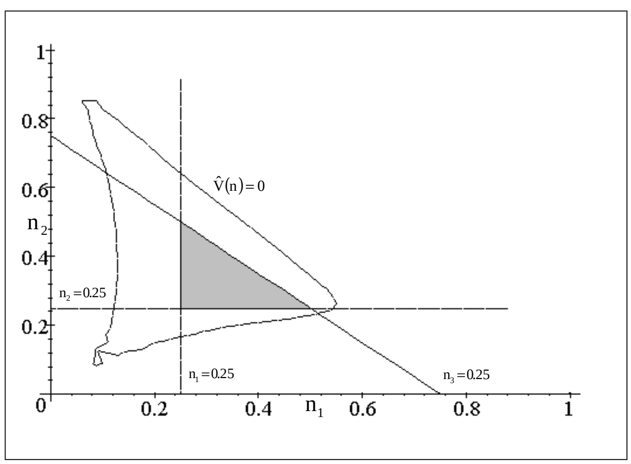


Figure 5.7: $\hat{V}(\mathbf{n})$ when $n_3 > 0.25$

We turn now to the stay or exit decision of player 2 given that player 3 stays in the game afterwards. Since $n_2 > n_2^2$, s/he stays in the game whenever player 1 has decided to stay in the game before. Suppose that player 1 has left the game and that $n_3 > 0.25$. Player 2 does not leave the game, because by taking $i = 2$ and $j = 3$ in Figure 5.6, we can prove that s/he gains from staying. Suppose now that $n_3 \leq 0.25$. Player 2 stays in the game if and only if the condition $n_2 (n_2 + n_3) \geq \left(\frac{n_2}{n_2+n_1}\right)^2$ is satisfied. We restate the weak inequality in the form $(1 - n_1)(1 - n_3)^2 \geq n_2 = 1 - n_1 - n_3$. This condition is equivalent to $(1 - n_1 - n_3 + n_1 n_3)(1 - n_3) \geq 1 - n_1 - n_3$. We perform all the necessary multiplications to arrive at $n_1 (1 - n_3) - n_3 (1 - n_1 - n_3) \geq 0$. Since $n_3^2 > 0$, we have shown that it is optimal for player 2 to stay in the game. Finally, player 1 decides to stay given that player 2 and 3 do not leave the game afterwards, because her/his payoff from leaving is $\delta n_1^2 < \delta n_1$. □

Proof of Proposition 5.2. By Lemma 5.5 we have that $s_t = 3$. Therefore, we can apply Lemma 5.4 to get that $U_l^* (\delta) = \delta n_l$ for all l which reduces to $U_l^* (\delta) = n_l$ as $\delta \to 1$. Moreover, the grand coalition forms independently of the Nature moves, because it has been seen in Lemmata 5.2 and 5.4 that every offer is accepted in equilibrium. □

Bibliography

[1] M. Alingham and A. Sandmo, *Income Tax Evasion: A Theoretical Analysis*, Journal of Public Economics **1** (1972), 323–338.

[2] K. Arrow, *Rational Choice Functions and Orderings*, Economica (1959), 121–127.

[3] N. Baigent and Y. Xu, *Independent Necessary and Sufficient Conditions for Approval Voting*, Mathematical Social Sciences **21** (1991), 21–29.

[4] K.H. Baik and S. Lee, *Collective Rent-Seeking with Endogeneous Group Sizes*, European Journal of Political Economy **13** (1997), 113–126.

[5] ______, *Strategic Groups and Rent Dissipation*, Economic Inquiry **39** (2001), 672–684.

[6] M. Barbie and C. Puppe, *Non-Manipulable Domains for the Borda Count*, Mimeo (2004).

[7] D. Berga and S. Serizawa, *Maximal Domain for Strategy-Proof Rules with One Public Good*, Journal of Economic Theory **90** (2000), 39–61.

[8] D. Blair and E. Muller, *Essential Aggregation Procedures on Restricted Domains of Preferences*, Journal of Economic Theory **30** (1983), 34–53.

[9] F. Bloch, *Sequential Formation of Coalitions in Games with Externalities and Fixed Payoff Division*, Games and Economic Behavior **14** (1996), 90–123.

[10] F. Bloch, S. Sánchez-Pagés, and R. Soubeyran, *When does Universal Peace Prevail? Secession and Group Formation in Rent Seeking Contests and Policy Conflicts*, forthcoming Economics of Governance.

[11] A. Blume, D. DeJong, Y. Kim, and G. Sprinkle, *Experimental Evidence on the Evolution of Meaning of Messages in Sender-Receiver Games*, American Economic Review **88** (1998), 1323–1339.

[12] A. Bogomolnaia, H. Moulin, and R. Stong, *Collective Choice under Dichotomous Preferences*, forthcoming Journal of Economic Theory.

[13] G. Bolton and A. Ockenfels, *ERC: A Theory of Equity, Reciprocity, and Competition*, American Economic Review **90** (2000), 166–193.

[14] J. Borda, *Mémoire sur les élections au Scrutin*, Histoire de l'Academie Royale des Sciences, Paris (1781).

[15] A. Bosch-Domènech, R. Nagel, J. García-Montalvo, and A. Satorra, *One, Two, (Three), ..., Infinity: Newspaper and Lab Beauty-Contest Experiments*, American Economic Review **92** (2002), 1687–1701.

[16] S. Brams and P. Fishburn, *Going from Theory to Practice: The Mixed Success of Approval Voting*, forthcoming Social Choice and Welfare.

[17] ———, *Approval Voting*, American Political Science Review **72** (1978), 831–847.

[18] S. Brams, P. Fishburn, and S. Merrill III, *The Responsiveness of Approval Voting: Comments on Saari and van Newenhizen*, Public Choice **59** (1988), 121–131.

[19] J. Brandts and G. Charness, *Truth or Consequence: An Experiment*, Management Science **49** (2003), 116–130.

[20] H. Cai and J. Wang, *Overcommunication and Bounded Rationality in Strategic Information Transmission Games: An Experimental Investigation*, Mimeo (2003).

[21] S. Ching and S. Serizawa, *A Maximal Domain for the Existence of Strategy-Proof Rules*, Journal of Economic Theory **78** (1998), 157–166.

[22] Marquis de Condorcet, *An Essay on the Application of Probability Decision Making: An Election between three Candidates (1785)*, in "The Political Theory of Condorcet" eds. F. Sommerlad and I. McLean, University of Oxford (1989).

[23] M. Costa-Gomes, V. Crawford, and B. Broseta, *Cognition and Behavior in Normal Form Games: An Experimental Study*, Econometrica **69** (2001), 1193–1235.

[24] V. Crawford, *Lying for Strategic Advantages: Rational and Boundedly Rational Misrepresentations of Intentions*, American Economic Review **93** (2003), 133–149.

[25] V. Crawford and J. Sobel, *Strategic Information Transmission*, Econometrica **50** (1982), 1431–1451.

[26] J. Dickhaut, K. McCabe, and A. Mukherji, *An Experimental Study of Strategic Information Transmission*, Economic Theory **6** (1995), 389–403.

[27] J. Duffy and N. Feltovich, *Do Actions Speak Louder Than Words? Observation vs. Cheap Talk as Coordination Devices*, Games and Economic Behavior **39** (2002), 1–27.

[28] ______, *Words, Deeds and Lies: Strategic Behavior in Games with Multiple Signals*, Mimeo (2003).

[29] M. Dummett, *The Borda Count and Agenda Manipulation*, Social Choice and Welfare **15** (1998), 289–296.

[30] J. Esteban and D. Ray, *Conflict and Distribution*, Journal of Economic Theory **87** (1999), 379–415.

[31] J. Esteban and J. Sákovics, *Olson vs. Coase: Coalitional Worth in Conflict*, Theory and Decision **55** (2004), 339–357.

[32] J. Farrell and M. Rabin, *Cheap Talk*, Journal of Economic Perspectives **10** (1996), 103–118.

[33] E. Fehr and K. Schmidt, *A Theory of Fairness, Competition, and Cooperation*, Quaterly Journal of Economics **114** (1999), 817–864.

[34] U. Fischbacher, *Z-Tree Tutorial Version 2.1*, Zurich University (2002).

[35] P. Fishburn, *Symmetric and Consistent Aggregation with Dichotomous Preferences*, in "Aggregation and Revelation of Preferences" (ed. J. Laffont), North-Holland, Amsterdam, 1978.

[36] ______, *Axioms for Approval Voting: Direct Proof*, Journal of Economic Theory **19** (1978), 180–185.

[37] E. Galor, *Information Sharing in Oligopoly*, Econometrica **53** (1985), 329–343.

[38] U. Gneezy, *Deception: The Role of Consequences*, forthcoming American Economic Review.

[39] F. Gul, *Bargaining Foundations of the Shapley Value*, Econometrica **57** (1989), 81–95.

[40] P. Heidhues and J. Lagerlof, *Hiding Information in Electoral Competition*, Games and Economic Behavior **42** (2003), 48–74.

[41] K. Inada, *A Note on the Simple Majority Decision Rule*, Econometrica **32** (1964), 525–531.

[42] Y. Luo, C. Yue, and T. Chen, *Strategy Stability and Sincerity in Approval Voting*, Social Choice and Welafre **13** (1996), 17–23.

[43] K. May, *A Set of Independent Necessary and Sufficient Conditions for Simple Majority Decision*, Econometrica **20** (1952), 680–684.

[44] R. McKelvey and T. Palfrey, *Quantal Response Equilibria in Extensive Form Games*, Experimental Economics **1** (1995), 9–41.

[45] ______, *Quantal Response Equilibria in Normal Form Games*, Games and Economic Behavior **10** (1998), 6–38.

[46] J. Morgan and P. Stocken, *An Analysis of Stock Recommendations*, RAND Journal of Economics **34** (2003), 183–203.

[47] H. Moulin, *Axioms of Cooperative Decision Making*, Cambridge University Press, 1988.

[48] D. Pérez-Castrillo and T. Verdier, *La structure industrielle dans une course au brevet avec coûts fixes et coûts variables*, Revue Economique **42** (1991), 1111–1140.

[49] M. Regenwetter and I. Tsetlin, *Approval Voting and Positional Voting Methods: Inference, Relationship, Examples*, Social Choice and Welfare **22** (2004), 539–566.

[50] A. Roth, T. Sönmez, and U. Ünver, *Pairwise Kidney Exchange*, Mimeo (2004).

[51] D. Saari, *Susceptibility to Manipulation*, Public Choice **61** (1990), 21–41.

[52] D. Saari and J. van Newenhizen, *Is Approval Voting an 'Unmitigated Evil'?: A Response to Brams, Fishburn, and Merill*, Public Choice **59** (1988), 132–147.

[53] ______, *The Problem of Indeterminacy in Approval, Multiple, and Truncated Voting Systems*, Public Choice **59** (1988), 101–120.

[54] R. Schmalensee, *A Model of Promotional Competition in Oligopoly*, Review of Economic Studies **43** (1976), 71–76.

[55] A. Sen, *Maximization and the Act of Choice*, Econometrica **65** (1997), 745–779.

[56] M. Sertel, *Characterizing Approval Voting*, Journal of Economic Theory **45** (1988), 207–211.

[57] L. Shapley and M. Shubik, *Trade Using one Commodity as a Means of Payment*, Journal of Political Economy **85** (1977), 937–967.

[58] D. Smith, *Manipulability Measures of Common Social Choice Functions*, Social Choice and Welfare **16** (1999), 639–661.

[59] G. Tan and R. Wang, *Endogeneous Coalition Formation in Rivalry*, Mimeo (1999).

[60] G. Tullock, *The Welfare Costs of Tariffs, Monopolies and Theft*, Western Economic Journal **5** (1967), 224–232.