Asymmetric Information, Partially Defined Games, Minimal Manipulability

door

Stefan Maus

Asymmetric Information,
Partially Defined Games,
Minimal Manipulability

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit Maastricht,
op gezag van Rector Magnificus,
Prof. mr. G.P.M.F. Mols,
volgens het besluit van het College van Decanen,
in het openbaar te verdedigen
op donderdag 25 november 2004 om 16.00 uur

door

Stefan Maus

Contents

III Minimal manipulability of social choice functions 71

5 Unanimity and Nondictatorship 73

6 Surjectivity and Anonymity 97

7 Unanimity and Anonymity 117

Preface

In February 2001 I started as a PhD student at the University of Maastricht, with an appointment for four years. This thesis contains the results of that period. The official title of my Meteor-Project is "Approximate Strategy-Proofness", but since I was appointed to that project only in November 2001 the first two parts of this thesis have little to do with that topic. I am very grateful to my supervisors Hans Peters and Ton Storcken for their support and for their belief in me. For the pleasant working atmosphere my thanks also go to my colleagues (at the department and elsewhere) as well.

Furthermore, I thank my family and friends for their support. They could not help me with the research itself, but they gave me the necessary distraction from it. The greatest thanks in that respect go to my girlfriend Katja and our two "Kleinlinge", Lisa and Susi, who were always there for me when I was at home.

Chapter 1

Introduction

This thesis addresses three different topics and is based on six papers. The first part (Chapters 2 and 3) concerns exchange economies with asymmetric information.

An exchange economy is a model for a market where agents meet and exchange their initial bundles of goods for other, more preferred bundles. Asymmetric information captures the idea that agents can have different information about the goods that they exchange. When the agents have met at the market the resulting allocation of goods depends on the initial allocation of goods, the preferences of the agents and their information. It can be expected that with this resulting allocation, no group of agents wants to meet again and exchange their goods amongst them once more. If that is the case they could have done so at the market in the first place. Game-theoretically the set of allocations that has this property is called the core (of the market game). The market fails, whenever there are no allocations with the desired properties, meaning in mathematical terms that the core is empty. When all agents have the same information it can be shown that under quite acceptable assumptions the core is not empty. When different groups of agents have access to different information this is no longer true. This is due to the fact that core allocations are based on the information accessible by the so called grand coalition, i.e. the coalition consisting of all agents. The following example may serve to illustrate this.

Think of a market consisting of three agents. Agent 1 has four apples and agents 2 and 3 have four peaches each. However, the apples that agent 1 has may have worms. Let us call the apples without worms good apples and the apples with worms bad apples. Agents 2 and 3 know that there is a 50% chance that the apples they are offered are good apples. There are no such doubts about the quality of the peaches. Agent 1 knows which kind of apples he has. Agents 1 and 2 know that they can trust each other. When agents 1, 2 and 3 meet, however, agent $2's$ confidence in agent 1 disappears since agent 3 does not trust agent 1, and even convinces agent 2 not to trust agent 1. Now, suppose that the bad apples have no worth for agents 2 and 3, and that agents 2 and 3 want at least three good apples in their allocation after exchange. Agent 1 would like to have some peaches and some apples. Now, with the information that is used in the coalition of agents 1, 2 and 3,

agents 2 and 3 expect to be fooled by agent 1 with probability 50%, so they expect to get at most $50\% * 4 = 2$ good apples from agent 1 on average. Hence, in the grand coalition no exchange will take place. In the subcoalition of agents 1 and 2 however, these two agents could exchange without quality doubts and achieve an allocation that they both prefer. Hence, in this example the core is empty.

In Chapter 2 a condition on the information that the agents have is given that ensures that the core of the exchange economy is nonempty. In particular the condition rules out "insider trading", because this possibility creates problems with the existence of core allocations. When smaller groups of agents have access to superior information they may prefer to trade only among themselves, thus destroying the existence of core allocations which require that all agents can participate in the exchange of goods.

In real markets goods are usually allocated through prices, but these have no use in the model described so far. If there are prices for the goods, everybody is assumed to buy the most preferred bundle of goods among those that he can afford. This yields the demand side of the market. The total supply is given by the initial bundles of all agents. Prices should level out in such a way that demand equals supply. In economic terms, markets "clear" or are in equilibrium, and a price that clears the market is called an equilibrium price. The associated area in economics is called General Equilibrium Theory. Again with symmetric information the existence of such equilibrium can be shown under acceptable assumptions; with asymmetric information, however, it need not exist (see e.g. Akerlof 1970). Moreover, it can be shown that under symmetric information equilibrium allocations are also core allocations. This is a nice result, since it says that when the market has cleared through an equilibrium price system there is no need for the agents to meet again and redistribute their goods by exchange. There are, however, still more core allocations than equilibria, so it could be argued that organizing the market through prices is efficient but not as flexible as exchange. Here general equilibrium theory provides another result that says that if the market is large the core shrinks to the set of equilibrium allocations. So, in sufficiently large markets, when it is also hard to imagine negotiations of exchange among all agents, equilibrium prices are a good way to organize the market. How these prices come about is another problem. The validity of these results, however, has thus far only been proven under symmetric information. Chapter 3 investigates their validity in the presence of asymmetric information. It turns out that under the assumption of "no insider trading" as in Chapter 2, along with some other conditions on the information and the initial wealth of the agents, more or less similar results also hold under asymmetric information.

The second part (Chapter 4) of the thesis deals with partially defined cooperative games. Think of a project that can be undertaken by some firms. They may choose to form coalitions that cooperate, i.e. join forces to undertake the project, for manifold reasons. Each such coalition has a different worth because of the firms' different possibilities. The question arises how to divide the benefits of cooperation. Equal division will usually not be an option, e.g., there may be big and small firms. In game theory a solution of this

problem is called a value function. A value function divides the worth of the coalition of all firms among the firms. One will demand certain properties of the value function, e.g. a firm that contributes more to any coalition than another firm should also receive a higher share. Value functions satisfying this property are called monotonic. Two other quite natural properties are anonymity and efficiency. They say that firms whose contributions are the same should receive the same share, and that the shares of all firms together equal the worth of the coalition of all firms (nothing is left over and no manna comes from heaven to be divided). Young (1985) shows that there is only one value function satisfying these three properties, which is the Shapley value (Shapley, 1953). A drawback of the Shapley value is that the worths of all coalitions have to be known to compute it. If you think of a cooperation like Star Alliance, that is a lot of information. So, one might think of a cooperation where the worths of some coalitions are not known for sure. Such a cooperation is modelled by a partially defined cooperative game. Chapter 4 extends some results about the Shapley value to partially defined cooperative games by defining the expected Shapley value. By doing so Chapter 4 also unifies and extends some value functions that have been considered for partially defined cooperative games: the b-weighted Shapley value, the reduced Shapley value and the normalized Shapley value (Willson, 1993 and Housman, 2001). These correspond to having certain estimates about the missing values of cooperation, which may be interpreted as the game that agents expect to be playing on average - this is called the expected game in Chapter 4. The expected Shapley value is then the Shapley value of this expected game, and it can be characterized by similar properties as the Shapley value. It is shown that the expected games that underly the b-weighted Shapley value, the reduced Shapley value and the normalized Shapley value may be rather inadequate in certain situations, e.g. when agents know they are playing a monotonic game. The expected Shapley value offers more flexibility.

The third part of the thesis (Chapters 5, 6 and 7) considers minimally manipulable social choice functions. Consider the problem of a group of individuals having to agree on an action out of a certain set of possibilities, or to select a candidate from a set of candidates. Each individual has his own ranking of the opportunities, and for each possible combination of such rankings among the individuals some decision must be taken. A well-known example is voting by plurality. Since ties are not allowed, assume that when some alternatives receive most of the votes, a previously appointed "president of the committee" decides. Suppose that there are three alternatives to choose from, say A, B and C. Furthermore there are three individuals in the committee, call them agents $1, 2$ and 3. Now agent 1 prefers A to B and B to C and since he is rational also A to C. In short we say that agent 1 has preference ABC. Assume that agents 2 and 3 have preferences BCA and CAB respectively. The committee agrees to use majority voting and that agent 1 is to act as a president of the committee. In that case A should be chosen, since all alternatives receive one vote and then the president decides. But note that agent 2 prefers C to A and if he votes for C instead of B, C has two votes and is chosen. So, by misrepresenting his true ranking BCA by CAB,

agent 2 can get an outcome that is better for him. In social choice theory such decision problems are described by social choice functions, and the particular social choice function described above would be called manipulable at the profile (ABC, BCA, CAB). Now, one might think that, by using a different way to decide, manipulation can be avoided. That is true, but it has been shown that when there are more than two alternatives the only way to decide without allowing for manipulation at some profile is by appointing a dictator[1] (Gibbard, 1973, and Satterthwaite, 1975), which is not desirable.

Think for example of a situation with two individuals and three alternatives A, B and C. There are $3! = 6$ different rankings of these alternatives: ABC, ACB, BAC, BCA, CAB and CBA. Hence, there are $36 = 6^2$ different combinations of these rankings among the two individuals. In the following table, agent 1's ranking is specified in the first cell of the rows and agent 2's ranking in the first cell of the columns. The entries in each of the 36 combinations of rankings correspond to the social choice function where agent 1 is a dictator.

2 :	ABC	ACB	BAC	BCA	CAB	CBA
1 : ABC	A	A	A	A	A	A
1 : ACB	A	A	A	A	A	A
1 : BAC	B	B	B	B	B	B
1 : BCA	B	B	B	B	B	B
1 : CAB	C	C	C	C	C	C
1 : CBA	C	C	C	C	C	C

At a given cell agent 2 can change the outcome to any other outcome in the same row, and agent 1 can change the outcome to any other outcome in the same column. Clearly, for the dictatorial situation depicted none of them can manipulate in this way. Gibbard's and Satterthwaite's result says that if there are cells with A, B and C in them, but not in a dictatorial way, then there is a cell where an agent can manipulate, i.e. agent 1 can find a better outcome for him in the same column, or agent 2 can find a better outcome for him in the same row. For example, if we have C instead of A in the (ABC, ABC)–cell then four profiles become manipulable, indicated by square brackets in the following table.

2 :	ABC	ACB	BAC	BCA	CAB	CBA
1 : ABC	[C]	A	A	[A]	[A]	[A]
1 : ACB	A	A	A	A	A	A
1 : BAC	B	B	B	B	B	B
1 : BCA	B	B	B	B	B	B
1 : CAB	C	C	C	C	C	C
1 : CBA	C	C	C	C	C	C

---

[1]A dictator is an individual whose most preferred outcome is always chosen, i.e. if the dictator has ranking ABC then A is chosen, regardless of the rankings of all other individuals.

Kelly (1988) has shown that the minimal number of manipulable profiles that such two agent, three alternative social choice functions have to admit if they are nondictatorial and can select any alternative is 2. Note that there are about 3^{36} such tables (in each of 36 cells one of 3 alternatives), so it is of course not possible to analyze all of them. Fristrup and Keiding (1998) have shown that for two agents and $m \geq 4$ alternatives the minimal number is $2m - 1$. Apart from some other special cases in Kelly (1988) for two or three agents and three or four alternatives these were the only minimal numbers known in the literature.

In Chapters 5, 6 and 7 we investigate this issue of minimal manipulability under some additional assumptions on the social choice functions.

Chapter 5 asks additionally to nondictatorship for unanimity, which means that an alternative is chosen if all agents agree that it is their most preferred outcome. For example, choosing B at the profile (ABC, ACB) violates unanimity. At such profiles A has to be chosen. It turns out that the minimally manipulable social choice functions under these assumptions are almost dictatorial - i.e. an agent is a dictator except at one profile. If that profile is carefully chosen such social choice functions have $(n-1)(\frac{m!}{2}-1)+1$ manipulable profiles if there are $n \geq 3$ agents. With the help of this result the three alternative case without unanimity can be solved as well. In this case the minimal number of manipulable profiles is equal to the number of agents n. For a description of the social choice functions attaining this minimum the reader is referred to Section 5.6.

Chapter 6 asks for anonymity but not for unanimity. Anonymity means that the names of the agents do not matter, so e.g. at the profile (ABC, CBA) the same choice is made as at the profile (CBA, ABC). Anonymity rules out dictatoriality since anonymity implies that then all agents have to be the dictator, a clear contradiction at any profile where their most preferred outcomes do not coincide. The minimal number of manipulable profiles is found to be $n\left(\frac{m!}{3}-1\right)(m-2)$. However that case is not completely solved: we have to assume that $n \geq m+1$ if $m = 4$ and that $n \geq m+2$ if $m \geq 5$. For a description of the social choice functions attaining these minima if $m \geq 4$ the reader is referred to Chapter 6. The case $m = 3$ is contained already in Chapter 5, since the minimally manipulable surjective and nondictatorial social choice functions in Section 5.6 are anonymous.

Chapter 7 asks for unanimity and anonymity. Only the case of three alternatives is analyzed. The minimal number of manipulable profiles is found to be $2^n - 2$. Some of the social choice functions attaining that lower bound have a nice interpretation if a status quo z among the alternatives exists, e.g. a current jurisdiction z about which there is consensus that it should only be altered to one of two new jurisdictions x and y if every voter prefers the new jurisdiction. Then, the least manipulable social choice functions that guarantee anonymity and unanimity of the outcome are as follows. Choose z in the situation where there is consensus that no change should be made. In the other situations choose an alternative which is unanimously preferred to z; if necessary, i.e. if there are two such alternatives, choose according to an arbitrary unanimous and anonymous social choice function over two alternatives.

Part I

Exchange economies with asymmetric information

Part I

Exchange economies with asymmetric information

Chapter 2

Balancedness and the core

A condition is given that is equivalent to balancedness of all NTU-games derived from an exchange economy with asymmetric information when endowments are variable. The condition is applicable to the ex-ante model with expected utilities, but also to the more general model of Arrow-Radner type economies without subjective probabilities. Differences in the interpretation of measurability assumptions between these two models are discussed, and another model with information consistent utility functions is developed in which the result would also hold.[1]

2.1 Introduction

The standard Arrow-Debreu model of an exchange economy can be extended to account for uncertainty by differentiating the commodities over the states of nature, as first proposed already by Debreu (1959) in Chapter 7 of his 'Theory of Value'. This approach has been extended to account for asymmetric information in a seminal paper by Radner (1968, 1982), using economies where agents possess different information. It is assumed, that an agent can carry out only such trades that are compatible with his information structure.

After this development in the fifties and sixties, the seventies and eighties saw a growing literature on Rational Expectations Equilibrium (REE), which is a natural extension of Arrow-Debreu's deterministic model of Walrasian equilibrium to a differential information framework. However, prices in a fully revealing REE are not able to reflect the informational asymmetries of agents. This left room for criticism and further work.

A new literature emerges from the work of Wilson (1978), who considers the core of an economy with differential information. Wilson considered the problem of how agents within a coalition share information. He gave two different scenarios, corresponding to the notion of fine core (pooling information) and coarse core (use only common information) of an economy with asymmetric information structure. He uses an interim core concept,

---

[1]The author thanks particularly Hans Peters for helpful suggestions. The results in this chapter were first formulated in Maus (2002).

where agents engage in coalitional negotiations after receiving their private information. For both core notions problems with existence and incentive compatibility arise, so that this approach was not pursued further for some time.

In the early nineties, the core was reconsidered. Yannelis (1991) considered a new information sharing concept based on measurability constraints, and thus introduced the private core. This core concept exists under quite general assumptions, and it is coalitional Bayesian incentive compatible, as shown in Koutsougeras and Yannelis (1993). Furthermore, it rewards the better informed agents, and it provides sensible outcomes in some situations where REE fails too exist. Allen (1991a, 1991b, 1994) carries the development further, by introducing more general information sharing rules. Recently, Einy et al. (2001) have shown a core equivalence theorem for large economies.

In one of her mentioned articles, Allen (1994) considers the ex-ante core, where agents have to form coalitions before the true state of the world is revealed to them. She allows for arbitrary communication systems rather than fine or coarse communication. This raises the question to find conditions on the communication system that assure nonemptiness of the ex-ante core. In this chapter such a condition is given. It turns out to be an equivalent to balancedness of all NTU-games derived from the exchange economy with asymmetric information when endowments are viewed as variable. Moreover, the condition is applicable to the model without expected utility, as introduced in Schwalbe (1999). Most articles consider agents' preferences derived from state-dependent preferences by taking expectations with respect to some subjective probability measure over the states of the world. However, such a description of preferences is not necessary. As Debreu (1959) explains in Chapter 7 of his 'Theory of Value', and as is the case in Radner (1968), preferences under uncertainty can be formulated without referring to probabilities. The case of expected utility functions is then included in this more general approach.

The analysis is organized as follows: In Section 2.2 the formal model is developed and the core concepts are defined. Section 2.3 contains the results and some examples to illustrate the relation of the new condition with other conditions. In Section 2.4 a discussion is included on measurability assumptions and how the implications of these differ for the models with and without expectations. Some concluding remarks follow in Section 2.5.

2.2   Preliminaries

In this section the relevant definitions for the notion of an exchange economy with asymmetric information are given. Two different alternatives are considered, one, where the agents have state-dependent utility functions and priors, and another one, where the utility function is defined on the state-commodity space altogether. The state-dependent utility case includes the expected utility model with finitely many states of the world as in Allen (1994). The case of a utility function avoiding the usage of expected utility by defining the utility function on the state-commodity space appears in Schwalbe (1999). There, balancedness

for all exchange economies with asymmetric information is shown. But this holds only if one assumes the feasible allocations for the economy to be determined on the maximum amount of information available, that is, the information an agent would have if all coalitions could be formed simultaneously. Here, I will only assume that agents can use the information available to them in the grand coalition. This seems less artificial, and is in agreement with the expected utility model of Allen (1994).

2.2.1 Information

Let Ω be the finite set of states of the world. Let $\mathcal{P}^*$ be the set of partitions of Ω. A $\mathcal{P} \in \mathcal{P}^*$ is called an *information set*. The interpretation is that states contained in an element $S \in \mathcal{P}$ cannot be distinguished under that information set. For each $\omega \in \Omega$ denote by $\mathcal{P}(\omega)$ the element of the partition $\mathcal{P}$ that contains ω.

An information set $\widetilde{\mathcal{P}}$ is called *finer* than $\mathcal{P}$, if every element of $\widetilde{\mathcal{P}}$ is contained in an element of $\mathcal{P}$. $\mathcal{P}$ is then called *coarser* than $\widetilde{\mathcal{P}}$.

Let N be a finite set of agents. Each agent has an initial endowment of information, described by $\mathcal{P}_i \in \mathcal{P}^*$. Then forming coalitions the information of agents may change, e.g. due to communication. Let $\mathcal{P}_i^S$ be the information that agent $i \in S$ has if the coalition S is formed. Throughout I assume that $\mathcal{P}_i^{\{i\}} = \mathcal{P}_i$. A collection $(\mathcal{P}_i^S)_{i \in S, S \subseteq N}$ is called a *communication system*.

$$\mathcal{P}_i^m := \bigvee_{S \ni i} \mathcal{P}_i^S := \{\bigcap_{S \ni i} \mathcal{P}_i^S(\omega) \mid \omega \in \Omega\}$$

is the maximum amount of information of agent i. Checking that $\mathcal{P}_i^m \in \mathcal{P}^*$ is straightforward.

An information set $\mathcal{P}$ generates a σ-algebra $\sigma(\mathcal{P})$. A communication system $(\mathcal{P}_i^S)_{i \in S, S \subseteq N}$ is called *nested* if $\sigma(\mathcal{P}_i^S) \subseteq \sigma(\mathcal{P}_i^T)$ or equivalently $\mathcal{P}_i^S \subseteq \sigma(\mathcal{P}_i^T)$ for all $i \in S \subseteq T$. It is called *bounded* if $\sigma(\mathcal{P}_i^S) \subseteq \sigma(\mathcal{P}_i^N)$ for all $i \in S \subseteq N$.

Information restricts the possible net trades of an agent. He cannot trade different amounts on events that he cannot distinguish. Formally this is captured by the following. Let $\mathcal{P}$ be the information the agent has. Then his trades of k goods are limited to the following set of functions

$$X_{\mathcal{P}} := \{x \mid x : \Omega \to \mathbb{R}^k \text{ and } x \text{ is } \sigma(\mathcal{P})\text{-measurable}\}.$$

Hence, $x \in X_{\mathcal{P}}$ if and only if x is constant on elements of $\mathcal{P}$. The characteristic function of the set Ω, denoted by

$$\mathbb{I}_\Omega : \Omega \to \mathbb{R}$$
$$: \omega \longmapsto 1,$$

is in $X_{\mathcal{P}}$ for every $\mathcal{P}$ in a one good economy for example. When there are $k > 1$ goods

$$\mathbb{I}_\Omega \underbrace{\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}}_{\in \mathbb{R}^k}$$

is in $X_{\mathcal{P}}$ for every $\mathcal{P}$. I will denote $X_{\mathcal{P}_i^S}$ by X_i^S and $X_{\mathcal{P}_i^m}$ by X_i^m. Then, if $S = \{i\}$ I will write X_i. Call $X_{\mathcal{P}}$ the set of *informational feasible* trades under $\mathcal{P}$, X_i^S the set of *informational feasible trades of agent i in coalition S* and $\prod_{i \in S} X_i^S$ the set of *informational feasible trades of the coalition S*.

2.2.2   Exchange economies with asymmetric information

An *exchange economy with asymmetric information* $\mathbf{E}$ is given by

1. a finite set of agents $N = \{1, 2, \ldots, n\}$,

2. a finite set Ω of states of the economy,

3. the initial endowments $e_i : \Omega \to \mathbb{R}^k$ for every agent $i \in N$,

4. the communication system $(\mathcal{P}_i^S)_{i \in N, S \subseteq N}$,

5. the utility functions $u_i : \mathbb{R}^{k \times \Omega} \to \mathbb{R}$ for every agent $i \in N$ or

5.' state-dependent utility functions $u_i' : \Omega \times \mathbb{R}^k \to \mathbb{R}$ and strictly positive subjective probabilities $\mu_i(\omega) > 0$.

When using 5 I will write

$$\mathbf{E} = (N, \Omega, (e_i, u_i, (\mathcal{P}_i^S)_{i \in S \subseteq N})_{i \in N})$$

and speak of the *model without expectations*, when using 5' I will write

$$\mathbf{E} = (N, \Omega, (e_i, u_i', \mu_i, (\mathcal{P}_i^S)_{i \in S \subseteq N})_{i \in N})$$

and speak of the *expected utility model*.

A vector of net trades $(z_i)_{i \in S}$ satisfying $e_i + z_i \geq 0$ for all $i \in S$ and $\sum_{i \in S} z_i = 0$ is called *physically feasible* for the coalition $S \subseteq N$.

Moreover, for the results in this chapter it is necessary to assume that

6. the utility functions are quasi-concave in $\mathbb{R}^{k \times \Omega}$

6.' the state-dependent utility functions $u(\omega, \cdot)$ are concave in $\mathbb{R}^k$.

By taking expectations

$$u_i(x_i) := E_{\mu_i} u_i'(\omega, x_i(\omega)) := \sum_{\omega \in \Omega} \mu_i(\omega) u_i'(\omega, x_i(\omega)),$$

it becomes clear, that the expected utility model with concave state-dependent utilities is really a subclass of the model without expectations. In Allen (1994) it was assumed, that the state-dependent utility functions be concave, which is reflected here in assumption 6'. The proofs in the more general case of assumptions 5 and 6 demand only quasi-concavity of the utility function. As the integral of concave functions is concave, this encompasses the case of state-dependent concave utilities. Nothing similar can be said about the case of state-dependent quasi-concave utilities, since the integral of quasi-concave functions need not be quasi-concave.

Here no assumptions are made with respect to measurability of initial endowments or state-dependent utility functions. It might be regarded as unreasonable, to have initial endowments or utility functions that contain more information than any of the information partitions the agent has in the coalitions of the game. For the decision to form coalitions and trade in these coalitions the information contained in endowments and utility seems essential, and therefore it might be argued that the agent has to know it. I will discuss these matters in Section 2.4.

2.2.3 The core

An *NTU-game in characteristic function form* is a correspondence $V : 2^N \setminus \{\emptyset\} \rightarrow \mathbb{R}^N$ satisfying

1. $V(S)$ is nonempty and closed for $S \neq \emptyset$,

2. if $x \in V(S)$ and $y \in \mathbb{R}^N$ is such that $y_i \leq x_i$ for all $i \in S$ then $y \in V(S)$,

3. for every $i \in N$ there is an $m_i \in \mathbb{R}$ with $V(\{i\}) = \{x \in \mathbb{R}^N \mid x_i \leq m_i\}$, and $V(N) \cap \{x \in \mathbb{R}^N \mid x_i \geq m_i \; \forall i \in N\}$ is nonempty and compact.

A collection of coalitions $B \subseteq 2^N \setminus \{\emptyset\}$ is *balanced* if there are positive real numbers λ_S for every $S \in B$ such that $\sum_{S \in B: i \in S} \lambda_S = 1$ for every $i \in N$. An NTU-game V is *balanced* if $\bigcap_{S \in B} V(S) \subseteq V(N)$ for every balanced collection B. Scarf (1967) has proved that if V is balanced, then it has a nonempty core.

The *NTU-game associated with the expected utility model* is defined by

$$V_E^{eu}(S) = \left\{ x \in \mathbb{R}^N \mid \text{there exists } (z_i)_{i \in S} \in \prod_{i \in S} X_i^S \text{ such that } e_i + z_i \geq 0, \right.$$

$$\left. \sum_{i \in S} z_i = 0 \text{ and } x_i \leq E_{\mu_i} u_i(\omega, (e_i + z_i)(\omega)) \right\}$$

for each coalition $S \neq \emptyset$.

Allen (1994) shows that this defines indeed an NTU-game. Let $\text{int}(X)$ denote the *interior* of a set $X \subseteq \mathbb{R}^N$ with respect to the usual topology on $\mathbb{R}^N$. The *expected utility core* of the exchange economy with asymmetric information is then defined to be the NTU-core of the associated NTU-game:

$$C^{eu}(\mathbf{E}) := C(V_{\mathbf{E}}^{eu}) = V_{\mathbf{E}}^{eu}(N) \setminus \bigcup_{\emptyset \neq S \subseteq N} \text{int}(V_{\mathbf{E}}^{eu}(S)).$$

The *NTU-game associated with the model without expectations* is defined by

$$V_{\mathbf{E}}(S) \;=\; \left\{ x \in \mathbb{R}^N \mid \text{there exists } (z_i)_{i \in S} \in \prod_{i \in S} X_i^S \text{ such that } e_i + z_i \geq 0, \right.$$

$$\left. \sum_{i \in S} z_i = 0 \text{ and } x_i \leq u_i(e_i + z_i) \right\}$$

for each coalition $S \neq \emptyset$.

The *core of the exchange economy with asymmetric information* is then defined to be the NTU-core of the associated NTU-game:

$$C(\mathbf{E}) := C(V_{\mathbf{E}}) = V_{\mathbf{E}}(N) \setminus \bigcup_{\emptyset \neq S \subseteq N} \text{int}(V_{\mathbf{E}}(S)).$$

2.3 Balancedness of the market games with asymmetric information

Schwalbe (1999) defines the *maximum information* of an agent i as the information $\mathcal{P}_i^m$ he could gain by joining all coalitions simultaneously. His set of *feasible allocations* E for the whole economy then consists of all physically feasible allocations $(x_i)_{i \in N} \in \prod_{i \in N} X_i^m$. That makes the set E large enough to assure balancedness of the associated market game with asymmetric information for any communication system. As there are no obvious reasons for using this notion of maximum information, I will give here another condition to assure balancedness based on the sets X_i^S, and use $\mathcal{P}_i^N$, rather than $\mathcal{P}_i^m$, for the whole economy. Note that $\mathcal{P}_i^m = \mathcal{P}_i^N$ if the communication system is bounded.

I give an example to show that the condition contained in Allen (1994), to guarantee (total) balancedness of the associated market game in the expected utility model, is not correct, and then present an alternative that is applicable to both models.

The following definitions are taken from Allen (1994).

Definition 1 *The communication system* $(\mathcal{P}_i^S)_{i,S}$ *is essentially nested if for all* $i \in N$ *and all coalitions* S *and* T *such that* $i \in S \subseteq T \subseteq N$, *if* $\Omega' \subset \Omega$ *is such that* $0 < \mu_i(\Omega') < 1$ *where* $\Omega' \in \sigma(\mathcal{P}_i^S)$, *and* $\Omega' = \Omega_1 \cup \Omega_2 \cup \ldots \cup \Omega_L, L \geq 1$ *for some disjoint* $\Omega_l \in \mathcal{P}_{j(l)}^S$ *with*

$j(l) \in S \setminus \{i\}$, $l = 1, \ldots, L$, then $\Omega' \in \sigma(\mathcal{P}_i^T)$ and each $\Omega_l \in \sigma(\mathcal{P}_{j(l)}^T)$. The communication system $(\mathcal{P}_i^S)_{i,S}$ is essentially bounded if the condition above holds for $T = N$.

Obviously essentially nested implies essentially bounded. Furthermore, nested implies essentially nested and bounded implies essentially bounded. Now the claim in Allen (1994) is, that essential nestedness is equivalent to total balancedness of the associated market games with varying utilities and endowments, and essential boundedness is equivalent to balancedness. I give a counterexample to show that this is not so. The game specified will be essentially nested, but not balanced, hence constituting a counterexample to both claims.

Example 2 Let there be five agents $N = \{1, 2, 3, 4, 5\}$, 4 states $\Omega = \{1, 2, 3, 4\}$ and prior $\mu_i(\omega) = \mu(\omega) = \frac{1}{4}$ for all $i \in N$. The communication system is given by

$$
\begin{aligned}
\mathcal{P}_1^S &= \{\{1, 2\}, \{3, 4\}\}, & \text{if } 1 \in S, |S| \leq 4, \\
\mathcal{P}_2^S &= \{\{1, 4\}, \{2, 3\}\}, & \text{if } 2 \in S, |S| \leq 4, \\
\mathcal{P}_3^S &= \{\{1, 3, 4\}, \{2\}\}, & \text{if } 3 \in S, |S| \leq 4, \\
\mathcal{P}_4^S &= \{\{1, 2, 3\}, \{4\}\}, & \text{if } 4 \in S, |S| \leq 4, \\
\mathcal{P}_5^S &= \{\Omega\}, & \text{if } 5 \in S, |S| \leq 4,
\end{aligned}
$$

and $\mathcal{P}_i^N = \{\Omega\}$ for all $i \in N$. This communication system is obviously essentially nested, as $\Omega' \subseteq \Omega$ with $0 < \mu_i(\Omega') < 1$, $\Omega' \in \sigma(\mathcal{P}_i^S)$, and $\Omega' = \Omega_1 \cup \Omega_2 \cup \ldots \cup \Omega_L$ for some disjoint $\Omega_l \in \mathcal{P}_{j(l)}^S$ with $j(l) \in S \setminus \{i\}$, $l = 1, \ldots, L$, does not exist for any $i \in S \subseteq N$.

Now consider the following one good economy

$$
\begin{aligned}
e_1 &= (1, 1, 0, 0) &, u_1 : \Omega \times \mathbb{R}, u(\omega, x) = \mathbb{I}_{\{1,2\}}(\omega)x + 3\mathbb{I}_{\{3,4\}}(\omega)x, \\
e_2 &= (0, 1, 1, 0) &, u_2 : \Omega \times \mathbb{R}, u(\omega, x) = \mathbb{I}_{\{2,3\}}(\omega)x + 3\mathbb{I}_{\{1,4\}}(\omega)x, \\
e_3 &= (0, 0, 0, 0) &, u_3 : \Omega \times \mathbb{R}, u(\omega, x) = x, \\
e_4 &= (0, 0, 0, 1) &, u_4 : \Omega \times \mathbb{R}, u(\omega, x) = x, \\
e_5 &= (1, 1, 1, 1) &, u_5 : \Omega \times \mathbb{R}, u(\omega, x) = x.
\end{aligned}
$$

There will be no exchange of goods in the grand coalition as only $\{\emptyset, \{\Omega\}\}$-measurable trades are allowed there and so only agent 5 could possibly trade as all others have a state with 0 endowment. So the unique candidate for the core is the utility vector $(u_i(e_i))_{i \in N}$ arising from the initial endowment e. The expected utility is then

$$
\begin{aligned}
u_1(e_1) &= \frac{1}{2} \\
u_2(e_2) &= \frac{1}{2}, \\
u_3(e_3) &= 0, \\
u_4(e_4) &= \frac{1}{4}, \\
u_5(e_5) &= 1.
\end{aligned}
$$

Consider the coalition $S = \{1, 2, 3, 4\}$. The net trades

$$z_1 = (-1, -1, \frac{1}{2}, \frac{1}{2}),$$

$$z_2 = (\frac{1}{2}, -1, -1, \frac{1}{2}),$$

$$z_3 = (0, 1\frac{1}{2}, 0, 0),$$

$$z_4 = (\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, -1)$$

are informational and physically feasible and lead to the following expected utility

$$u_1(e_1 + z_1) = \frac{3}{4} > u_1(e_1) = \frac{1}{2},$$

$$u_2(e_2 + z_2) = \frac{3}{4} > u_2(e_2) = \frac{1}{2},$$

$$u_3(e_3 + z_3) = \frac{3}{8} > u_3(e_3) = 0,$$

$$u_4(e_4 + z_4) = \frac{3}{8} > u_4(e_4) = \frac{1}{4},$$

and hence there is a deviation to the utility vector $(u_i(e_i))_{i \in N}$. So the core is empty and the game cannot be balanced.

Remark 3 *The restriction of information in the grand coalition to $\{\Omega\}$ might seem strong. The essential idea is only that net trades that are possible in a subcoalition are removed in the grand coalition and hence deviations of this subcoalition become possible. Thus, the assumption $\mathcal{P}_i^N = \{\Omega\}$ is only made for simplicity.*

The idea behind the condition of essential nestedness/boundedness was that the possibility of net trades is not only tied to physical feasibility, but also to informational feasibility. Hence, not all combinations of physically feasible net trades have to be considered as possible deviations. I propose the following definition that captures this idea better than essentially nestedness/boundedness.

Definition 4 *The communication system $(\mathcal{P}_i^S)_{i,S}$ is trade nested if for all coalitions $S \subseteq T \subseteq N$ and all $(z_j)_{j \in S} \in \prod_{j \in S} X_j^S$ with $\sum_{j \in S} z_j = 0$, it holds that $(z_j)_{j \in S} \in \prod_{j \in S} X_j^T$. The communication system $(\mathcal{P}_i^S)_{i,S}$ is trade bounded if the condition above holds for $T = N$.*

Clearly, the system in Example 2 is neither trade nested nor trade bounded. Now I claim the following for the model without expectations.

Theorem 5 *Fix the number of goods k and the finite sets Ω and N arbitrarily. Consider all exchange economies with asymmetric information and with these parameters fixed. As endowments and utilities vary, all NTU market games with asymmetric information are totally balanced if and only if $(\mathcal{P}_i^S)_{i,S}$ is trade nested.*

Proof. Necessity: Suppose that the communication system $(\mathcal{P}_i^S)_{i,S}$ is not trade nested. Then one can construct an exchange economy with asymmetric information that is not totally balanced, as follows. By assumption there are coalitions $S \subset T$ and $(z_j)_{j \in S} \in \prod_{j \in S} X_j^S$ such that $\sum_{j \in S} z_j = 0$ and $(z_j)_{j \in S} \notin \prod_{j \in S} X_j^T$. Consequently there is an agent $\hat{j} \in S$ such that $z_{\hat{j}} \notin X_{\hat{j}}^T$. Now let K be large enough to ensure that

$$\tilde{z}_j := z_j - K \mathbb{I}_\Omega \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} < 0$$

for all $j \in S, j \neq \hat{j}$. Then put $e_j := -\tilde{z}_j$ for all $j \in S, j \neq \hat{j}$ and $e_j := 0$ for all $j \in N \setminus S$ and for $j = \hat{j}$. Let the utility functions of the agents $j \neq \hat{j}$ be the zero-functions and for agent $\hat{j}$ let $u_{\hat{j}}(x) := \frac{1}{|\Omega|} \sum_{\omega \in \Omega} \sum_{m=1}^k x_m(\omega)$. I show that the subgame (T, V) is not balanced. Consider the balanced collection $\mathcal{B} := \{S, T \setminus S\}$ with constant weight function $\mathbb{I}_\mathcal{B}$. Obviously

$$V(T \setminus S) = \{x \in \mathbb{R}^T \mid x_i \leq 0 \text{ for all } i \in T \setminus S\}.$$

The best that $\hat{j}$ can get in coalition S is the sum of all the initial endowments of the others

$$\tilde{z}_{\hat{j}} = \sum_{j \in S, j \neq \hat{j}} e_j = - \sum_{j \in S, j \neq \hat{j}} \tilde{z}_j = - \sum_{j \in S, j \neq \hat{j}} z_j + (|S| - 1)K \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \mathbb{I}_\Omega$$

$$= z_{\hat{j}} + (|S| - 1)K \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \mathbb{I}_\Omega > 0.$$

So $\tilde{z}_{\hat{j}}$ is $\mathcal{P}_{\hat{j}}^S$ measurable and

$$u_{\hat{j}}(\tilde{z}_{\hat{j}}) = \frac{1}{|\Omega|} \sum_{\omega \in \Omega} \sum_{m=1}^k \tilde{z}_{\hat{j}m}(\omega) > 0.$$

As u_j is the zero function for $j \neq \hat{j}$ we have

$$V(S) = \{x \in \mathbb{R}^T \mid x_i \leq 0 \text{ for all } i \in S \setminus \{\hat{j}\} \text{ and } x_{\hat{j}} \leq u_{\hat{j}}(\tilde{z}_{\hat{j}})\}.$$

Hence

$$\bigcap_{B \in \mathcal{B}} V(B) = \{x \in \mathbb{R}^T \mid x_i \leq 0 \text{ for all } i \in T \setminus \{\hat{j}\} \text{ and } x_{\hat{j}} \leq u_{\hat{j}}(\tilde{z}_{\hat{j}})\}.$$

But on the other hand, as $(z_j)_{j \in S} \notin \prod_{j \in S} X_j^T$ one has also that $(\tilde{z}_j)_{j \in S} \notin \prod_{j \in S} X_j^T$. Furthermore, as $\tilde{z}_{\hat{j}} = e(T) := \sum_{j \in T} e_j$ and $e(T \setminus S) = 0$, any physically feasible vector of net trades $(z_j')_{j \in T} \in \prod_{j \in T} X_j^T$ leads to a utility for agent $\hat{j}$ that is strictly less than $u_{\hat{j}}(\tilde{z}_{\hat{j}})$. So the balancedness condition

$$V(T) \supseteq \bigcap_{B \in \mathcal{B}} V(B)$$

is violated in the subgame (T, V). This shows that the game (N, V) is not totally balanced.

Sufficiency: To show total balancedness of the derived game (N, V), it has to be shown that every subgame (T, V), $T \subseteq N$ is balanced, i.e. $V(T) \supseteq \bigcap_{B \in \mathcal{B}} V(B)$ holds for every balanced collection $\mathcal{B}$ of subsets of T. So let $\mathcal{B}$ be an arbitrary balanced collection of subsets of T with weights $\gamma : \mathcal{B} \to (0, 1]$. Let $x \in \bigcap_{B \in \mathcal{B}} V(B)$. For every $B \in \mathcal{B}$ one has $x \in V(B)$, and hence there are by definition of $V(B)$ net trades $z_i^B, i \in B$, that are informational and physically feasible in the coalition B such that $x_i \leq u_i(e_i + z_i^B)$ for all $i \in B$. Extend $(z^B)_{i \in B}$ to a vector of net trades in the grand coalition T of the subgame by putting $z_i^B := 0$ for all $i \in T \setminus B$. The communication system is trade nested so $(z_i^B)_{i \in T} \in \prod_{i \in B} X_i^T$, and as $0 \in X_i^T$ this implies $(z_i^B)_{i \in T} \in \prod_{i \in T} X_i^T$. Moreover $\sum_{i \in T} z_i^B = \sum_{i \in B} z_i^B = 0$. Consequently $(z_i^B)_{i \in T}$ is an informational and physically feasible vector of net trades in T. Now let $z_i := \sum_{B \in \mathcal{B}, B \ni i} \gamma(B) z_i^B = \sum_{B \in \mathcal{B}} \gamma(B) z_i^B$ for every $i \in T$. As it follows easily that the sets X_i^T are convex and contain the z_i^B for every $B \in \mathcal{B}$ one obtains that also $z_i \in X_i^T$, i.e. $(z_i)_{i \in T}$ is informational feasible in the coalition T. It is clear that $e_i + z_i \geq 0$ for all $i \in T$ so

$$\sum_{i \in T} z_i = \sum_{i \in T} \sum_{B \in \mathcal{B}} \gamma(B) z_i^B = \sum_{B \in \mathcal{B}} \gamma(B) \sum_{i \in T} z_i^B$$
$$= \sum_{B \in \mathcal{B}} \gamma(B) \sum_{i \in B} z_i^B = 0$$

shows that $(z_i)_{i \in T}$ is also physically feasible. Quasiconcavity of the utility functions now implies that for all $i \in T$

$$u_i(e_i + z_i) = u_i(e_i + \sum_{B \in \mathcal{B}} \gamma(B) z_i^B) = u_i\left(\sum_{B \in \mathcal{B}, B \ni i} \gamma(B) e_i + \sum_{B \in \mathcal{B}, B \ni i} \gamma(B) z_i^B \right)$$
$$= u_i\left(\sum_{B \in \mathcal{B}, B \ni i} \gamma(B)(e_i + z_i^B) \right) \geq \min_{B \in \mathcal{B}, B \ni i} \{ u_i(e_i + z_i^B) \}$$
$$\geq x_i,$$

so $x \in V(T)$, and that shows $V(T) \supseteq \bigcap_{B \in \mathcal{B}} V(B)$ for all subgames (T, V) and balanced collections $\mathcal{B}$ of subsets of T. Hence the derived game is totally balanced. ∎

Corollary 6 *Fix the number of goods k and the finite sets Ω and N arbitrarily. Consider all exchange economies with asymmetric information and with these parameters fixed. As endowments and utilities vary, all NTU market games with asymmetric information are balanced if and only if $(\mathcal{P}_i^S)_{i,S}$ is trade bounded.*

Proof. For necessity and sufficiency set $T = N$ in the corresponding parts of the proof of Theorem 5. ∎

Corollary 7 *Fix the number of goods k and the finite sets Ω and N arbitrarily. Consider all exchange economies with asymmetric information and with these parameters fixed. As endowments and utilities vary, all NTU market games with asymmetric information are*

(i) balanced if and only if $(\mathcal{P}_i^S)_{i,S}$ is trade bounded and

(ii) totally balanced if and only if $(\mathcal{P}_i^S)_{i,S}$ is trade nested.

Proof. The expected utility case is a subclass of the other model recovered by taking $u_i : \mathbf{R}^{k \times \Omega} \to \mathbf{R}$ to be the expected utility functions

$$\sum_{\omega \in \Omega} \mu_i(\omega) u_i(\omega, x(\omega))$$

for all $i \in N$. ∎

As balancedness implies that the core of the NTU-game is not empty and total balancedness implies that the core of all subgames is not empty one obtains another corollary.

Corollary 8 *Fix the number of goods k and the finite sets Ω and N arbitrarily. Consider all exchange economies with asymmetric information and with these parameters fixed. As endowments and utilities vary, a sufficient condition for all exchange economies with asymmetric information to have a non-empty core is trade boundedness of the communication system. A sufficient condition for all subgames to have a non-empty core is trade nestedness of the communication system.*

Note that the game constructed in the proof of Theorem 5 to show that nestedness is also a necessary condition has a non-empty core. In fact zero utility for every agent, corresponding to no trade, is in the core. So the corollary cannot state equivalence of nonemptiness of the core (subcores) and boundedness (nestedness).

Obviously, nestedness (boundedness) of the communication system implies trade nestedness (boundedness).

To see that the conditions of trade nestedness and trade boundedness are really weaker than nestedness and boundedness of the communication system, consider the following example.

Example 9 *Let there be three agents $N = \{1, 2, 3\}$ and 4 states $\Omega = \{1, 2, 3, 4\}$. The communication system given by*

$\mathcal{P}_1^S = \{\{1\}, \{2\}, \{3, 4\}\}, \quad \text{if } 1 \in S, |S| \leq 2,$
$\mathcal{P}_2^S = \{\{1, 2, 3\}, \{4\}\}, \quad \text{if } 2 \in S,$
$\mathcal{P}_3^S = \{\{1, 2, 4\}\{3\}\}, \quad \text{if } 3 \in S,$
$\mathcal{P}_1^N = \{\{1, 2\}, \{3, 4\}\},$

is trade nested but not nested. This is easily seen, as any physically feasible trades that agent 1 can make must necessarily be constant on the set $\{1, 2\}$.

On the other hand trade nestedness (boundedness) is stronger than essentially nestedness (boundedness) in the expected utility model.

Theorem 10 *Any trade nested (bounded) communication rule is essentially nested (bounded).*

Proof. Let the communication system $(\mathcal{P}_i^S)_{i,S}$ be trade nested. Fix arbitrary i, S and T such that $i \in S \subseteq T \subseteq N$. Suppose there is $\Omega' \subset \Omega$ such that $0 < \mu_i(\Omega') < 1$ where $\Omega' \in \sigma(\mathcal{P}_i^S)$, and $\Omega' = \Omega_1 \cup \Omega_2 \cup \ldots \cup \Omega_L$ for some disjoint $\Omega_l \in \mathcal{P}_{j(l)}^S$ with $j(l) \in S \setminus \{i\}$, $l = 1, \ldots, L$. If such sets Ω', Ω_l do not exist for any combination of i, S and T the communication system is by definition trivially essentially nested. If they exist put $z_i = -\mathbb{I}_{\Omega'}, z_{j(l)} = \mathbb{I}_{\Omega_l}$ and $z_j = 0$ for all

$$ j \in S \setminus (\{i\} \cup \{j(l) \mid l = 1 \ldots L\}). $$

Then $(z_j)_{j \in S} \in \prod_{j \in S} X_j^S$ and $\sum_{j \in S} z_j = 0$ hence by trade nestedness $(z_j)_{j \in S} \in \prod_{j \in S} X_j^T$. But this implies that $\Omega' \in \sigma(\mathcal{P}_i^T)$ and each $\Omega_l \in \sigma(\mathcal{P}_{j(l)}^T)$. As i, S and T were arbitrary this shows that the communication system is essentially nested. The boundedness case follows by specialising T to the grand coalition N. ∎

2.4   Discussion of Measurability assumptions

As far as the initial endowment is concerned, an assumption such that every information partition $\mathcal{P}_i^S, i \in S \subseteq N$ is finer than the information contained in the initial endowment e_i, i.e. the coarsest partition $\mathcal{P}$ that makes e_i $\sigma(\mathcal{P})$−measurable, seems reasonable. After all, an agent must know his initial endowment for planning his net trade, hence he can use that information in every coalition. Observation of the initial endowments by the agents would not be necessary anymore if one would use utility functions defined on the net trades z rather than the commodity bundle $e_i + z$, and assign infinitely low utility $-\infty$ to net trades where $e_i + z \notin \mathbb{R}_+^{k \times \Omega}$. In detail, one switches in the general model from $u_i : \mathbb{R}^{k \times \Omega} \to \mathbb{R}$ to

$$ \tilde{u}_i : \mathbb{R}^{k \times \Omega} \to \mathbb{R} $$
$$: z \longmapsto \begin{cases} u_i(e_i + z) & \text{if } e_i + z \geq 0, \\ -\infty & \text{otherwise,} \end{cases} $$

and in the expected utility model from $u_i : \Omega \times \mathbb{R}^k \to \mathbb{R}$ to

$$ \tilde{u}_i : \Omega \times \mathbb{R}^k \to \mathbb{R} $$
$$: (\omega, z) \longmapsto \begin{cases} u_i(\omega, e_i(\omega) + z) & \text{if } e_i(\omega) + z \geq 0 \\ -\infty & \text{otherwise.} \end{cases} $$

Note that quasiconcavity of utility functions is not disturbed by this switch. However the "penalty restriction",

$$ e_i + z \notin \mathbb{R}_+^{k \times \Omega} \Rightarrow \tilde{u}_i(e_i + z) = -\infty, $$

on the utility functions that can be used is imposed. Now an agent plans his net trade and observes and outcome in $\mathbb{R}$ or $\mathbb{R}^\Omega$, depending on which model is used. Assume an agent observes an outcome of $-\infty$ in the model without expectations. He could then infer that

$e_i + z \geq 0$ is violated. But as he can only trade constant on states he cannot discern he will not be able to find out which state(s) lead to infinitely low utility, and therefore he can gain no information that he does not already have.

Unfortunately, in the expected utility model the agent observes a vector of outcomes, and if these differ he can still gain information. Measurability of the outcome vector $(\widetilde{u}_i(\omega, z(\omega)))_{\omega \in \Omega} \in \mathbb{R}^{\Omega}$ plays also a role here. As decisions in this chapter are based on the expected utility of that outcome vector one can easily withdraw from a thorough a discussion of these matters by opting for the usage of the model without expectations to circumvent these problems. But in general for state-dependent utility it might be useful to use state-dependent utility functions that depend also on the information partition the agent is using. That makes sense if the agent has to know his utility function to decide. Utility at some state ω and information partition $\mathcal{P}$ must then equal utility in the nondiscernable states $\omega' \in \mathcal{P}(\omega)$. A definition of a state-dependent utility function would then be

$$\widetilde{u} : \Omega \times \mathbb{R}^{k \times \Omega} \times \mathcal{P}^* \to \mathbb{R},$$

such that the outcome vector $(\widetilde{u}(\omega, x, \mathcal{P}))_{\omega \in \Omega}$ is $\mathcal{P}$–measurable for all $x \in \mathbb{R}^{k \times \Omega}$ that are informational feasible, i.e. $x - e_i$ is $\mathcal{P}$–measurable. In the expected utility model given here this could be reached by letting the utility function $\widetilde{u}$ at $\mathcal{P} \in \mathcal{P}^*$ be a version of the conditional expected utility of the state-dependent utility function $u : \Omega \times \mathbb{R}^k \to \mathbb{R}$. To be precise, let

$$\widetilde{u}(\omega, x, \mathcal{P}) = \frac{1}{\mu(\mathcal{P}(\omega))} \sum_{\omega' \in \mathcal{P}(\omega)} \mu(\omega') u(\omega', x(\omega')).$$

Moreover, it might be argued that subjective probabilities should also reflect information. That could easily be incorporated by taking information dependent subjective probabilities $\widetilde{\mu}(\cdot, \mathcal{P})$, that are consistent in the sense that they arise from an underlying probability measure μ independent of the information partition, i.e.

$$\widetilde{\mu} : \Omega \times \mathcal{P}^* \to \mathbb{R}$$
$$: (\omega, \mathcal{P}) \longmapsto \frac{\mu(\mathcal{P}(\omega))}{|\mathcal{P}(\omega)|}.$$

Obviously $\widetilde{u}$ and $\widetilde{\mu}$ at a particular $\mathcal{P}$ reveal no more information than is contained in $\mathcal{P}$ already. Furthermore, this replacement does not change the expected utility function in

the expected utility model, as may be seen from

$$\sum_{\omega \in \Omega} \tilde{\mu}(\omega, \mathcal{P}) \tilde{u}(\omega, x(\omega), \mathcal{P}) = \sum_{\omega \in \Omega} \frac{1}{|\mathcal{P}(\omega)|} \sum_{\omega' \in \mathcal{P}(\omega)} \mu(\omega') u(\omega', x(\omega'))$$

$$= \sum_{\omega \in \Omega} \sum_{\omega' \in \mathcal{P}(\omega)} \frac{1}{|\mathcal{P}(\omega')|} \mu(\omega') u(\omega', x(\omega'))$$

$$= \sum_{S \in \mathcal{P}} |S| \sum_{\omega \in S} \frac{1}{|S|} \mu(\omega) u(\omega, x(\omega))$$

$$= \sum_{\omega \in \Omega} \mu(\omega) u(\omega, x(\omega)).$$

Summarizing, agent i now plans in a particular coalition S a net trade z that is $\mathcal{P}_i^S$–measurable, observes a utility outcome $(\tilde{u}(\omega, e_i + z, \mathcal{P}_i^S))_{\omega \in \Omega}$ that is also $\mathcal{P}_i^S$–measurable, and uses that outcome to take expectation under an also $\mathcal{P}_i^S$–measurable subjective probability $\mu(\cdot, \mathcal{P}_i^S)$. So the expected utility model of Allen (1994) can be rewritten, with an additional restriction on the used utility functions, to reveal no information to the agent at the planning stage.

All results in this chapter would remain valid, if one would transform an exchange economy modelled without or with expectations in the way suggested in this section. What can be said in general for an exchange economy arising from this new framework, but not having an underlying representation in the model with or without expectations, is left open here.

Another approach would be to view the information contained in $\mathcal{P}_i^S$ as the information one is allowed to use for trading, and any additional information contained in the initial endowment or utility vectors is information that is e.g. insider information, and hence trading based on this information is forbidden. Page (1997) interprets the boundedness restriction on a communication system into that direction, calling it 'no insider trading'. Pooling information may be impossible to certain degrees as well which would lead to coarser information in larger coalitions.

2.5   Conclusion

In this chapter a condition on the communication system was given to assure balancedness of all NTU-games derived from an asymmetric exchange economy, when endowments are variable. The NTU-games can be derived in an ex-ante setting without using state-dependent utility and subjective probabilities. The condition remains valid in the ex-ante expected utility core. It provides an alternative to the condition given in Allen (1994). Possible extensions are to consider trade boundedness in an interim core concept or to study implications for value or equilibrium concepts. Various results are already available for special communication systems leading to the weak fine, fine and private core. Einy et al. (2001), for example, show that with a continuum of traders and 'irreducibility' of the

economy the set of equilibrium allocations in the sense of Radner (1968, 1982) coincides with the private core of Yannelis (1991). Moreover, they show that such an equilibrium allocation exists.

The existence of private and fine value allocations in an ex-ante sense was established in Krasa and Yannelis (1994, 1996). They also argued that there are problems with the coarse value. Again one might consider arbitrary communication systems and investigate whether trade boundedness leads to an existence result for the value.

Chapter 3

Competitive equilibrium and the core

A replica theorem is shown to hold for exchange economies with asymmetric information. In a replicated exchange economy with asymmetric information the set of all core elements with equal treatment is nonempty, but it is in general only a subset of the core. Nevertheless, the replica theorem and the presence of at least one core element with equal treatment suffice to show existence of a competitive quasi-equilibrium. Conditions on the initial endowments and the communication system are given to ensure that every competitive quasi-equilibrium is a competitive equilibrium.[1]

3.1 Introduction

The purpose of this chapter is to study the relationship between the core and competitive allocations in economies with asymmetric information. This extends the analysis of uncertainty, as introduced in Chapter 7 of Debreu's 'Theory of Value' (1959). There, uncertainty is incorporated into the Arrow-Debreu theory by introducing a finite set of states of the world, and viewing the commodities as differentiated by state. Still, every agent is assumed to possess the same full information set, i.e. there are no states that cannot be distinguished. This analysis is extended by Radner (1968, 1982), to cover the case of private information. Every agent is assigned a partition of the set of states of the world, with the interpretation that the sets in the partition contain those states which cannot be distinguished by the agent. A trade of an agent then has to be compatible with his information, that is, he cannot act differently on states that he cannot distinguish, or mathematically, his net trades have to be measurable with respect to the σ-algebra generated by his information partition. Radner (1968) extends the notion of Arrow-Debreu competitive equilibrium to this model. Trades are contingent contracts before information about the realized state of nature is obtained. Information cannot be communicated and restricts the budget sets of the agents. A Radner competitive equilibrium rewards the better informed trader in the

---

sense that his budget set is larger, thus he will in general be better off.

From the core perspective, the incorporation of information structures into the core dates back to Wilson's seminal paper (1978). However, with the core, there is more to be considered than just private information. One can devise an information structure to be parameterized not only by the agent space but also by the coalitions that an agent can be in. Private information would then be the case, where an agent's information does not change, regardless of the coalition he belongs to. But a lot of other cases are imaginable, e.g. agents pooling their information, leading to a situation in which superset coalitions have more information and the grand coalition has full information. Such a fine information system, and others, as the coarse, strong coarse and weak fine information system, have been the subject of various papers, investigating the existence of core allocations. As a starting point to this literature one can consider, apart from Wilson's work of course, the papers of Yannelis (1991) and Allen (1994). The frameworks given there have to be carefully distinguished under the aspect that, in Yannelis (1991), an allocation is preferred to another by an agent if its expected utility over every single set in the information partition (mathematically the conditional expected utility vector) is greater. This is often referred to as an interim concept. Allen (1994) on the contrary considers an ex-ante concept, where an allocation is preferred to another if the expected utility is greater. Of course if an allocation is preferred in Yannelis (1991), it will be preferred also in Allen (1994), but not vice versa, making the ex-ante core a subset of the interim core. Note, however, that as pointed out yet again in the inspiring work of Debreu's Theory of Value (1959), the usage of (conditional) expected utility, requiring a probability measure over the states of the world and state-dependent utilities, is not essential. Instead one can consider a utility function which maps the state-commodity space to the reals. This function could be the outcome of taking expectation over states, but this only shows that this model is able to incorporate the expected utility case. Recently, the existence of ex-ante core allocations in this general context has been addressed by Maus (2002, see Chapter 2), providing a condition on the information partitions, which is equivalent to nonemptiness of the core with asymmetric information, when endowments are variables. This result will be used here to investigate the question of equal treatment in the core and the existence of an adapted version of the Radner competitive equilibrium. The core reflects the veto power of any coalition. The recent paper by Hervés-Beloso et al. (2003) characterizes Radner Equilibrium by using the veto power of the grand coalition only. They show that an allocation is a Radner equilibrium if and only if it can not be blocked by the grand coalition in every economy obtained by perturbing the original endowments in the direction of the equilibrium candidate.

The list of contributions on the relationship between Radner equilibrium and ex-ante core concept with asymmetric information comprises two other recent papers by Einy et al. (2000, 2001) working with a continuum of agents. In Einy et al. (2001) they show that in an ex-ante model, if the economy is 'irreducible', then a Radner equilibrium exists and the set of Radner competitive equilibrium allocations coincides with the private core. It is also

shown, that the weak fine core corresponds to competitive equilibrium of an economy, where the symmetric information of the agents is their pooled private information. In the other paper Einy et al. (2000) they show that in an interim model the fine core is a subset of the ex-post core, and consequently every fine core allocation is a selection from the equilibrium correspondence of the associated family of full information economies. Differences with the work presented in this chapter are manifold. First of all the core is defined here without referring to expected utility and prior beliefs. Secondly the concept is strictly ex-ante. Thirdly the set of agents is finite, though agents are replicated to get to the core convergence result. Fourth, the communication system is not specified. Last but not least, no free disposability of commodities is assumed. As pointed out already in a footnote by Radner (1982), p.945, the disposed commodities may not be measurable with respect to the information of any single trader, making free disposability objectionable. In this chapter, resource feasibility will be considered as denoting strict equality of the sum of the allocated commodities and the sum of the initial endowments. This complicates the existence of an equilibrium considerably, since existence results from the theory of production economies such as employed in Radner (1982) or Einy et al. (2001) cannot be used. The reason for that is that the proofs of these existence results rely on the assumption that production sets have a nonempty relative interior. In a pure exchange economy that assumption translates into disposability of at least part of the endowments. Therefore, these existence results are not applicable, and a number of other, partly new results is necessary. These results are the existence result for the core from Maus (2002, see Chapter 2), a result extending equal treatment of agents of the same type in replicated economies to the case of the core with asymmetric information, and a result on the connection between quasi-competitive Radner equilibria and the cores of the replicated economies. All of these results are also of interest on their own.

The chapter is organized as follows. In Section 3.2 the formal model is described along with the definition of core and competitive (quasi-)equilibrium. Section 3.3 discusses the relationship between competitive (quasi-)equilibrium allocations and core allocations in large economies. To arrive at an existence result for equilibrium, further investigation of 'equal treatment in the core' is necessary, which is dealt with in Section 3.4. Then Section 3.5 establishes existence of an competitive quasi-equilibrium, and introduces some conditions under which this turns into an existence result for competitive equilibrium.

3.2 Preliminaries

Throughout this chapter, $\mathbb{R}_+$ is the set $\{x \in \mathbb{R} \mid x \geq 0\}$, and $x > y$ for vectors or matrices means that every entry in x strictly exceeds the corresponding entry in y. When a constant $c \in \mathbb{R}$ is written in a place where a matrix or vector is expected, it is understood to be the matrix or vector where all entries have the value c. A product of matrices or vectors px in the same space is understood to be the sum of the products of the matching entries, i.e.

the scalar product (matrices have to be interpreted as vectors for this).

3.2.1 Information

Let Ω be the finite set of states of the world. Let $\mathcal{P}^*$ be the set of partitions of Ω. A $\mathcal{P} \in \mathcal{P}^*$ is called an *information set*. The interpretation is that states contained in an element $S \in \mathcal{P}$ cannot be distinguished under that information set. For each $\omega \in \Omega$ denote by $\mathcal{P}(\omega)$ the element of the partition $\mathcal{P}$ that contains ω.

Let N be a finite set of agents. Each agent has an initial endowment of information, described by $\mathcal{P}_i \in \mathcal{P}$. When forming coalitions the information of agents may change, e.g. due to communication. Let $\mathcal{P}_i^S$ be the information that agent $i \in S$ has if the coalition S is formed. Throughout I assume that $\mathcal{P}_i^{\{i\}} = \mathcal{P}_i$. A collection $(\mathcal{P}_i^S)_{i \in S, S \subseteq N}$ is called a *communication system*.

A communication system is called *private* if $\mathcal{P}_i^S = \mathcal{P}_i$ for all $S \ni i$, i.e. the information does not change.

An information set $\mathcal{P}$ generates a σ-algebra $\sigma(\mathcal{P})$.

Information restricts the possible net trades of an agent. He cannot trade different amounts on events that he cannot distinguish. Formally this is captured by the following. Let $\mathcal{P}$ be the information the agent has. Then his trades of k goods are limited to the following set of functions

$$X_{\mathcal{P}} := \{x \mid x : \Omega \to \mathbb{R}^k \text{ and } x \text{ is } \sigma(\mathcal{P})\text{-measurable}\}.$$

Hence, $x \in X_{\mathcal{P}}$ if and only if x is constant on elements of $\mathcal{P}$. Thus, $x : \Omega \to \mathbb{R}^k$ can be identified with $x : \mathcal{P} \to \mathbb{R}^k$, where $x(A)$ is the constant value on A. The characteristic function of any set $B \in \mathcal{P}$, denoted by

$$\mathbb{I}_B : \Omega \to \mathbb{R}^k,$$

$$: \omega \longmapsto \begin{cases} 1, & \text{if } \omega \in B, \\ 0, & \text{if } \omega \notin B, \end{cases}$$

is in $X_{\mathcal{P}}$ for every $\mathcal{P}$ for example. I will denote $X_{\mathcal{P}_i^S}$ by X_i^S and if $S = \{i\}$ I will write X_i. Call $X_{\mathcal{P}}$ the set of *informational feasible trades under* $\mathcal{P}$, X_i^S the set of *informational feasible trades of agent i in coalition S* and $\prod_{i \in S} X_i^S$ the set of *informational feasible trades of the coalition S*.

Definition 11 *The communication system* $(\mathcal{P}_i^S)_{i,S}$ *is trade bounded if for all coalitions* $S \subseteq N$ *and all* $(z_j)_{j \in S} \in \prod_{j \in S} X_j^S$ *with* $\sum_{j \in S} z_j = 0$, *it holds that* $(z_j)_{j \in S} \in \prod_{j \in S} X_j^N$.

Informally speaking, this assures that every net trade possible in a subcoalition remains possible in the grand coalition.

3.2.2 Exchange economies with asymmetric information

An *exchange economy with asymmetric information* $\mathbf{E}$ is given by

1. a finite set of agents N,

2. a finite set Ω of states of the economy,

3. the initial endowments $e_i : \Omega \to \mathbf{R}_+^k$ for every agent $i \in N$,

4. the communication system $(\mathcal{P}_i^S)_{i \in S, S \subseteq N}$,

5. and utility functions $u_i : \mathbf{R}_+^{k \times \Omega} \to \mathbf{R}$ for every agent $i \in N$.

A vector of net trades $(z_i)_{i \in S}$ satisfying $e_i + z_i \geq 0$ for all $i \in S$ and $\sum_{i \in S} z_i = 0$ is called *physically feasible* for the coalition $S \subseteq N$. An *allocation* for a coalition S in an economy $\mathbf{E}$ is a function $x : S \to \mathbf{R}^{k \times \Omega}$ such that the net trades $x_i - e_i$ are informational and physically feasible for this coalition. In this chapter vectors $x \in \mathbf{R}_+^{k \times \Omega}$ are identified in the natural way with the space of functions $x : \Omega \to \mathbf{R}_+^k$. Furthermore, the following spaces of utility functions will be relevant:

$$
\begin{aligned}
U^{mo} \quad &:= \quad \{u : \mathbf{R}_+^{k \times \Omega} \to \mathbf{R} \mid u \text{ is strictly increasing,} \\
&\qquad \text{i.e. } y \geq x, y \neq x \Rightarrow u(y) > u(x)\}, \\
U_{qc} \quad &:= \quad \{u : \mathbf{R}_+^{k \times \Omega} \to \mathbf{R} \mid u \text{ is quasiconcave}\}, \\
U_{co,0}^{mo} \quad &:= \quad \{u : \mathbf{R}_+^{k \times \Omega} \to \mathbf{R} \mid u \text{ is concave and } u(0) = 0\} \cap U^{mo}, \\
U_{qc}^{mo} \quad &:= \quad U_{qc} \cap U^{mo}, \text{ and} \\
\mathbf{C}(\mathbf{R}_+^{k \times \Omega}) \quad &:= \quad \{u : \mathbf{R}_+^{k \times \Omega} \to \mathbf{R} \mid u \text{ is continuous}\}.
\end{aligned}
$$

3.2.3 The core

An *NTU-game in characteristic function form* is a correspondence $V : 2^N \setminus \{\emptyset\} \to \mathbf{R}^N$ satisfying

1. $V(S)$ is nonempty and closed for $S \neq \emptyset$,

2. if $x \in V(S)$ and $y \in \mathbf{R}^N$ is such that $y_i \leq x_i$ for all $i \in S$ then $y \in V(S)$,

3. for every $i \in N$ there is an $m_i \in \mathbf{R}$ with $V(\{i\}) = \{x \in \mathbf{R}^N \mid x_i \leq m_i\}$, and $V(N) \cap \{x \in \mathbf{R}^N \mid x_i \geq m_i \ \forall i \in N\}$ is nonempty and compact.

The *NTU-game associated with an exchange economy* $\mathbf{E}$ is defined by

$$
V_{\mathbf{E}}(S) \quad = \quad \Big\{ x \in \mathbf{R}^N \mid \text{there exists } (z_i)_{i \in S} \in \prod_{i \in S} X_i^S \text{ such that } e_i + z_i \geq 0,
$$

$$
\sum_{i \in S} z_i = 0 \text{ and } x_i \leq u_i(e_i + z_i) \Big\}
$$

for each coalition $S \neq \emptyset$. The core of this NTU-game is given by

$$C(V_{\mathbb{E}}) := V_{\mathbb{E}}(N) \backslash \bigcup_{\emptyset \neq S \subseteq N} \text{int}(V_{\mathbb{E}}(S)).$$

The core of the exchange economy with asymmetric information consists of all allocations for the grand coalition, resulting in a utility vector in the core of the associated NTU-game, i.e.

$$C(\mathbb{E}) := \{x \in \mathbb{R}^{N \times k \times \Omega} \mid x \text{ is an allocation for } N \text{ and } (u_i(x_i))_{i \in N} \in C(V_{\mathbb{E}})\}.$$

So $C(\mathbb{E})$ comprises all allocations x for the grand coalition to which no coalition $S \subseteq N$ has a *deviation*, i.e. an allocation y^S for that coalition such that $u_i(y_i^S) > u_i(x)$ for all $i \in S$.[2]

The following existence result for the core from Maus (2002, see Chapter 2) will be of use.

Theorem 12 *Let* $\mathbb{E}$ *be an exchange economy with asymmetric information and trade bounded communication system. Assume that* $u_i \in U_{qc}$ *for all* $i \in N$. *Then the NTU-core of* $V_{\mathbb{E}}$ *is not empty.*

3.2.4   Competitive Allocations

The information that an agent can use in a competitive allocation is assumed to be $\mathcal{P}_i^N$. This assumption could be viewed as implying that in a competitive equilibrium allocation the same communication takes places as in the grand coalition. Relations between core allocations and competitive allocations can in general not be expected if the measurability constraints for core allocations and competitive allocations are different.

A *price system* is a function $p : \Omega \to \mathbb{R}^k$, $p \neq 0$. Agents make contracts for the delivery of contingent commodities as in Chapter 7 of Debreu's Theory of Value (1959), i.e. before the state of the world is revealed to them. Payments are made for contingent delivery and irrevocable even if another state of the world is realized. Additionally every agent's trades are restricted by his information.

A price system contains the information

$$\sigma(p) := \bigcap \{\mathcal{A} \mid \mathcal{A} \; \sigma\text{-algebra on } \Omega, \; p \; \mathcal{A}\text{-measurable}\}.$$

It could be argued that agents i such that $\sigma(\mathcal{P}_i^N) \neq \sigma(p)$ can observe information from prices posted to them. This leads to rational expectations equilibrium, see for example Allen (1984), Radner (1979). We assume that agents observe private price signals, which then reveal no information to them. This is achieved by the following construction. Let p be a price system. For an information partition $\mathcal{P}$ the price system that can be *seen* under $\mathcal{P}$

---

[2]In the previous chapter this was defined to be just the set of utility vectors given by $C(V_{\mathbb{E}})$. Here we have to consider allocations instead of utility vectors because of the way in which we are going to relate the core and the set of equilibrium allocations.

is given by $(p(A))_{A \in \mathcal{P}}$, where $p(A) := \sum_{\omega \in A} p(\omega)$. Every agent observes (possibly different) futures prices, namely $(p(A))_{A \in \mathcal{P}_i^N}$, corresponding to his possible trading activities. Another way to view this, is to consider average prices $\tilde{p}$, which are different per agent, and are given by

$$\tilde{p}(\omega) := \frac{1}{|\mathcal{P}(\omega)|} \sum_{\tilde{\omega} \in \mathcal{P}(\omega)} p(\tilde{\omega}).$$

Under both viewpoints an agent having information $\mathcal{P}$ cannot gain information from the prices seen, as these are $\sigma(\mathcal{P})$-measurable. Furthermore, any informational feasible trading activity $z \in X^{\mathcal{P}}$ incurs the same cost under both viewpoints, as can be seen from

$$
\begin{aligned}
z\tilde{p} &= \sum_{\omega \in \Omega} z(\omega)\tilde{p}(\omega) \\
&= \sum_{A \in \mathcal{P}} \sum_{\omega \in A} z(\omega) \underbrace{\frac{1}{|\mathcal{P}(\omega)|} \sum_{\tilde{\omega} \in \mathcal{P}(\omega)} p(\tilde{\omega})}_{=A} \\
&= \sum_{A \in \mathcal{P}} z(A) \sum_{\omega \in A} \frac{1}{|A|} \sum_{\tilde{\omega} \in A} p(\tilde{\omega}) \\
&= \sum_{A \in \mathcal{P}} z(A) \sum_{\tilde{\omega} \in A} p(\tilde{\omega}) \\
&= \sum z(A)p(A).
\end{aligned}
$$

Note that if we assume that the price system p is set by a Walrasian auctioneer who knows the information that agents have, then we can assign also the task of computing the private price signals for the agents to him. The least information that the Walrasian auctioneer has to have then is the partitions $\mathcal{P}_i^N$ for any agent $i \in N$. Thus the information $\mathcal{P}^W$ that the auctioneer has will satisfy $\sigma(\mathcal{P}^W) \supseteq \sigma(\bigcup_{i \in N} \mathcal{P}_i^N)$ and this information constraints price vectors set by the auctioneer to be $\sigma(\mathcal{P}^W)$-measurable. However, as we shall see, this constraint holds automatically for the price systems that we construct in our proofs. These correspond to separating hyperplanes of $\sigma(\bigcup_{i \in N} \mathcal{P}_i^N)$-measurable sets of allocations, so they can be chosen to be $\sigma(\bigcup_{i \in N} \mathcal{P}_i^N)$-measurable themselves. In the end $\sigma(\bigcup_{i \in N} \mathcal{P}_i^N)$-measurability of price systems is used in Lemma 27 and in Corollary 28, but anywhere else in this chapter it is not required. Another question, that arises if we assume that the auctioneer receives signals $\mathcal{P}_i^N$ from the agents is, whether agents, either individually or coalitional, can manipulate the price the Walrasian auctioneer chooses by transmitting another information partition $\overline{\mathcal{P}}_i^N$ such that $\sigma(\overline{\mathcal{P}}_i^N) \subseteq \sigma(\mathcal{P}_i^N)$ to the auctioneer. This question is not addressed here. So, when thinking of prices as set by an auctioneer, it is assumed implicitly that either agents are honest or their information can be verified costless by the auctioneer.

Example 13 *Imagine that there are three states of the world $\Omega = \{1, 2, 3\}$, which indicate the quality of a commodity (say 1 is good, 2 is medium,3 is bad). Now, compare two agents having information partitions $\{\{1, 2\}, \{3\}\}$ and $\{\{1\}, \{2, 3\}\}$ respectively (the first agent can*

screen for bad quality and the second agent for good quality). Assume that a price system for that commodity is given by $p = (3, 2, 1)$. Then the first agent observes the price vector $(5, 1)$, where 5 is the price for contingent delivery of a unit amount in state 1 and 2, and the second agent observers the price vector $(3, 3)$, where 3 is the price for contingent delivery of a unit amount in states 2 and 3. If both agents were interested in buying only good quality, that would be cheaper for agent 2, as agent 1 has to buy the medium quality bundled with the good quality. Instead, if the good and the medium quality were appreciated similarly by the agents, the advantage would be on the side of agent 1. The average price systems here are given by $(2\frac{1}{2}, 2\frac{1}{2}, 1)$ and $(3, 1\frac{1}{2}, 1\frac{1}{2})$ for the first and the second agent respectively.◇

For a price system p define the *budget set* of an agent $i \in N$ by

$$B_i(p) = \{x \in \mathbb{R}_+^{k \times \Omega} \mid x - e_i \in X_i^N \text{ and } \sum_{A \in P_i^N} \sum_{w \in A} p(\omega)(x(\omega) - e_i(w)) \le 0\}$$

$$= \{z + e_i \in \mathbb{R}_+^{k \times \Omega} \mid z \in X_i^N \text{ and } \sum_{A \in P_i^N} p(A)z(A) \le 0\}$$

The budget set contains all contingent allocations that the agent can afford under the given price system and that result from informational feasible net trades. Now the definition of a competitive equilibrium can be stated.

Definition 14 *A competitive equilibrium for an economy* $\mathbb{E}$ *is a pair* (p, x), *where* $p \neq 0$ *is a price system and* x *is an allocation such that* x_i *maximizes* u_i *on* $B_i(p)$ *for all* $i \in N$. *A competitive allocation is an allocation* x, *for which there exists a price system* p *such that* (p, x) *is a competitive equilibrium.*

As mentioned in the introduction of the chapter, the model considered here is able to incorporate models which use an expected utility function derived from integration of state-dependent utilities with respect to a prior on states. This can be used to argue that Definition 14 covers the notion of Radner equilibrium (Radner, 1968) for pure exchange economies with private information, and that Corollary 28 is also an existence result for Radner equilibrium. A standard assumption made when considering Radner equilibrium, is that initial endowments are measurable with respect to the information of the agents in the grand coalition, i.e. that $e_i \in X_i^N$ for all $i \in N$. The reason that this is not assumed here is that some of the results, e.g. about the relationships between core and equilibrium, do not need this assumption. However, in the existence result of Corollary 28, there is the condition that the initial endowments of an agent should be measurable with respect to the information that the agent uses in any coalition that he is in, i.e. $e_i \in \bigcap_{S \ni i} X_i^S$. Letting information be private, i.e. $X_i^S = X_i^N = X_i^{\{i\}}$ for all $i \in N$ and $S \ni i$, this is just the private measurability of initial endowments assumed in Radner equilibrium. Hence, Corollary 28 turns into an existence result for Radner equilibrium in pure exchange economies with private information that meet the assumptions of the

corollary. This demands that expected utility as derived from taking expectations in the Radner setting is quasiconcave, strictly monotone and continuous. This is met for example if the prior is strictly positive, and state-dependent utilities are concave, monotone and continuous. This is usually assumed when dealing with existence of Radner equilibrium, e.g. already in the original paper of Radner (1968).

Next the weaker notion of competitive quasi-equilibrium is defined.

Definition 15 *A competitive quasi-equilibrium for an economy* $\mathbf{E}$ *is a pair* (p, x) *where* $p \neq 0$ *is a price system and* x *is an allocation such that* x_i *maximizes* u_i *on* $B_i(p)$ *whenever* $\inf p X_i^{N,\geq 0} := \inf\{px \mid x \in X_i^N, x \geq 0\} < pe_i$.

Obviously, every competitive equilibrium is also a competitive quasi-equilibrium. There is no difference between these two equilibrium notions if $\inf p X_i^{N,\geq 0} < pe_i$ for all $i \in N$. When utility functions are strictly increasing, these definitions imply that $(p(A))_{A \in \mathcal{P}^N} \geq 0$ in any competitive quasi-equilibrium and $(p(A))_{A \in \mathcal{P}^N} > 0$ in any competitive equilibrium. Hence, in that case one has de facto $0 = \inf p X_i^{N,\geq 0}$ and the condition $\inf p X_i^{N,\geq 0} < pe_i$ could be replaced by $0 < pe_i$ (positive income).

3.3 Competitive and core allocations in large economies

In this section sufficient conditions on the communication system are given such that the replica theorem holds. So first of all the replica economies $\mathbf{E}^n$, $n \in \mathbb{N}$, are defined.

Definition 16 *Let* $\mathbf{E} = (N, \Omega, (e_i, u_i, (\mathcal{P}_i^S)_{i \in S, S \subseteq N})_{i \in N})$ *be an exchange economy with asymmetric information. The* $n - th$ *replica*

$$\mathbf{E}^n = (N^n, \Omega, (\tilde{e}_i, \tilde{u}_i, (\tilde{\mathcal{P}}_i^S)_{i \in S, S \subseteq N^n})_{i \in N^n}), n \in \mathbb{N},$$

is the exchange economy with asymmetric information, where the set of agents is $N^n := N \times \{1, \ldots, n\}$ *and for an agent* $(i, j) \in N^n$ *one puts*

1. $\tilde{e}_{(i,j)} := e_i$

2. $\tilde{u}_{(i,j)} := u_i$

3. $\tilde{\mathcal{P}}_{(i,j)}^S := \mathcal{P}_i^{pr^1(S)}$, *where* pr^1 *is the projection on the first coordinate.*

So the information that agents can use in a coalition in the replica economy depends only on the type of the agents in that coalition and not on the total number of agents of a type in the coalition. Denote the agents of a given type $i \in N$, which are present in a coalition $S \subseteq N^n$, by

$$pr_i(S) := \{j \mid 1 \leq j \leq n \text{ and } (i, j) \in S\}.$$

Theorem 17 *Let* $\mathbf{E}$ *be an exchange economy with asymmetric information. Assume that* $e_i \in X_i^N$ *and* $u_i \in U^{mo} \cap C(\mathbb{R}_+^{k \times \Omega})$ *for all* $i \in N$. *If* $x \in \mathbb{R}^{N \times k \times \Omega}$ *is such that* $x^n = \prod_{j=1}^n x \in C(\mathbf{E}^n)$ *for all* $n \in \mathbb{N}$, *then* x *is a competitive quasi-equilibrium allocation. Moreover the price system* p *decentralizing* x *can be chosen* $\sigma(\bigcup_{i \in N} P_i^N)$-*measurable.*

Proof. Note that if the initial endowments e_i are P_i^N−measurable, then not only the net trades $(z_i)_{i \in N}$ leading to a competitive equilibrium or core allocation are P_i^N−measurable, but also the final allocation $(e_i + z_i)_{i \in N}$. Let x be such that $x^n \in C(\mathbf{E}^n)$ for all $n \in \mathbb{N}$. It has to be shown, that there is a price system $p \in \mathbb{R}^{k \times \Omega}$ such that for any agent $i \in N$ either $\inf pX_i^{N, \geq 0} = pe_i$, or x_i maximizes u_i on $B_i(p)$.

$$H(i) := (\{y \in \mathbb{R}_+^{k \times \Omega} \mid u(y) > u(x_i)\} - \{e_i\}) \cap X_i^N$$

is the set of all net trades z, where agent $i \in N$ prefers $e_i + z$ to x_i, and which are informational feasible for the agent in the grand coalition. Denote by $co(A)$ the convex hull of $A \subseteq \mathbb{R}^{k \times \Omega}$ and by $\mathbb{I}_B$ the characteristic function of $B \subseteq \Omega$. Let

$$X_- := \{z \in \mathbb{R}^{k \times \Omega} \mid z = -\sum_{i \in N} \sum_{P \in P_i^N} \alpha_{i,P} \mathbb{I}_P, \ \alpha_{i,P} > 0\}.$$

I claim that

$$co(\bigcup_{i \in N} H(i)) \cap X_- = \emptyset.$$

Suppose on the contrary that there is some $z \in co(\bigcup_{i \in N} H(i)) \cap X_-$, then

$$z = \sum_{a \in A} \lambda_a y_a, \text{ where } A \subseteq N, \ \sum_{a \in A} \lambda_a = 1, \ \lambda_a > 0, \ y_a \in H(a), \text{ and}$$

$$z = -\sum_{i \in N} \sum_{P \in P_i^N} \alpha_{i,P} \mathbb{I}_P, \ \alpha_{i,P} > 0.$$

It suffices to construct a contradiction to $x^n \in C(\mathbf{E}^n)$ for all $n \in \mathbb{N}$ for the case where the λ_a are rational. In this case let n be so large that $r_a := n\lambda_a \leq n$ is a natural number for every $a \in A$. Define an allocation x^* for the coalition

$$S := \{(a, j) \mid 1 \leq j \leq r_a \text{ and } a \in A\} \cup \{(i, n+1) \mid i \in N\}$$

in the economy $\mathbf{E}^{n+1}$ by

$$x_{(a,i)}^* := e_a + y_a \ (1 \leq i \leq r_a, a \in A) \text{ and}$$

$$x_{(i,n+1)}^* := x_i + n \sum_{P \in P_i^N} \alpha_{i,P} \mathbb{I}_P \ (i \in N).$$

This allocation is informational feasible for S, as $\widetilde{P}_{(i,j)}^S := P_i^{pr^{-1}(S)} = P_i^N$. Every member of S prefers his bundle in x^* to that in x^{n+1}. Furthermore, x^* is also physically feasible for S,

as

$$\sum_{i \in S} x_i^* = \sum_{a \in A} \sum_{1 \le i \le r_a} x_{(a,i)}^* + \sum_{i \in N} x_{(i,n+1)}^*$$

$$= \sum_{a \in A} (r_a e_a + r_a y_a) + \sum_{i \in N} (x_i + n \sum_{P \in \mathcal{P}_i^N} \alpha_{i,P} \mathbb{1}_P)$$

$$= \sum_{a \in A} r_a e_a + n \underbrace{\sum_{a \in A} \lambda_a y_a}_{=z} + \underbrace{x(N)}_{=e(N)} + n \underbrace{\sum_{i \in N} \sum_{P \in \mathcal{P}_i^N} \alpha_{i,P} \mathbb{1}_P}_{=-z}$$

$$= \sum_{a \in A} r_a e_a + e(N) + nz - nz$$

$$= e(S).$$

Therefore, x^* is a deviation in $\mathbf{E}^{n+1}$ of the coalition S to the allocation $x^{n+1} \in C(\mathbf{E}^{n+1})$, a contradiction.

Now, as

$$co(\bigcup_{i \in N} H(i)) \cap X_- = \emptyset,$$

$$co(\bigcup_{i \in N} H(i)) \text{ is convex and}$$

$$X_- \text{ is open and convex,}$$

there is by a version of the separation theorem for convex sets, a $p \in \mathbb{R}^{k \times \Omega}, p \ne 0$, such that

$$pz \ge 0 \text{ if } z \in co(\bigcup_{i \in N} H(i))$$

$$\text{and } pz \le 0 \text{ if } z \in X_-.$$

This p can be chosen to be $\sigma(\cup_{i \in N} \mathcal{P}_i)$-measurable, as any $z \in co(\cup_{i \in N} H(i)) \cup X_-$ will per se, viewed as a function from Ω to $\mathbb{R}^k$, be $\sigma(\cup_{i \in N} \mathcal{P}_i)$-measurable. To show that (x, p) is a competitive quasi-equilibrium, I will need that the price system $(p(A))_{A \in \mathcal{P}_i^N}$, that can be seen by any agent under his information partition $\mathcal{P}_i^N$ in the grand coalition, is nonnegative. Assume to the contrary that there is an $A \in \mathcal{P}_{\tilde{i}}^N$ for some $\tilde{i} \in N$ such that $p(A) < 0$. Let $z \in X_-$ be given by $\alpha_{\tilde{i},A} = K$ and $\alpha_{i,P} = 1$ otherwise. Then

$$pz = p(-\sum_{i \in N} \sum_{P \in \mathcal{P}_i^N} \alpha_{i,P} \mathbb{1}_P) = -\sum_{i \in N} \sum_{P \in \mathcal{P}_i^N} \alpha_{i,P} (p \mathbb{1}_P)$$

$$= -\sum_{i \in N} \sum_{P \in \mathcal{P}_i^N} \alpha_{i,P} p_P = -(\sum_{i \in N} \sum_{P \in \mathcal{P}_i^N, i \ne \tilde{i} \lor P \ne A} p_P) - K p(A)$$

$$= C - K p(A)$$

where the constant C is independent of K, and $p(A) < 0$. Therefore choosing K larger than $\frac{C}{p(A)}$ will make pz positive, contradicting $z \in X_-$. So $(p(A))_{A \in \mathcal{P}_i^N} \ge 0$ for any $i \in N$,

implying that $\inf pX_i^{N,\geq 0} = 0$ for any $i \in N$. It remains to be shown that x_i maximizes u_i on $B_i(p)$ when $0 = \inf pX_i^{N,\geq 0} < pe_i$. So suppose there is a $y \in B_i(p)$ such that $u_i(y) > u_i(x_i)$. Then $y - e_i \in H(i) \subseteq co(\cup_{i \in N} H(i))$, which together with $y \in B_i(p)$ implies that $py = pe_i$. As the income of the agent is positive, y is $\mathcal{P}_i^N$–measurable, and $p(A) \geq 0$ for all $A \in \mathcal{P}_i^N$, it can be concluded from

$$0 < pe_i = py = \sum_{\omega \in \Omega} p(\omega)y_i(\omega) = \sum_{A \in \mathcal{P}_i^N} \sum_{\omega \in A} p(\omega)y_i(\omega) = \sum_{A \in \mathcal{P}_i^N} p(A)y_i(A)$$

that there must be some $A \in \mathcal{P}_i^N$ such that $y_i(A) > 0$ and $p(A) > 0$. So, by lowering consumption equally in the states contained in A, a nonnegative sequence of y_n converging to y can be obtained such that $py_n < pe_i$ for all $n \in N$, i.e. $y_n - e_i \notin H(i)$. But then $u_i(y_n) \leq u_i(x_i)$, as y_n is not in $H(i)$. Since the utility functions u_i are assumed to be continuous this implies $u_i(y) \leq u_i(x_i)$, a contradiction to $u_i(y) > u_i(x_i)$. So x_i indeed maximizes u_i on $B_i(p)$ when agents have positive income. This concludes the proof that (x, p) is a quasi-equilibrium of the economy E. ∎

The next example shows that the $\mathcal{P}_i^N$-measurability of the initial endowments cannot be dropped from the assumptions. The underlying reason is that an agent cannot use all of his initial endowment when trading, due to the measurability restrictions on net trades. However, when his income is calculated that is done by valuing the whole initial endowment, leading to higher demand than what can actually be achieved by trading. Thus, one could circumvent these restrictions of $e_i \in X_i^N$ by defining the income of the agents in another way, taking into account only the parts of the initial endowments which can really be used for trading. When allocations have to be nonnegative as in this chapter, this would be endowments e'_i such that $e'_{ij}(\omega) := \min_{\omega' \in \mathcal{P}_i(\omega)} e_{ij}(\omega')$ for all $\omega \in \Omega$ and commodities j.

Example 18 *Consider a private information economy with 3 agents $i \in N = \{1, 2, 3\}$, one commodity and three states $\Omega = \{1, 2, 3\}$. Let the initial endowments, the (private) information and the utility functions be*

$$e_1 := (2, 1, 2), \ \mathcal{P}_1 := \{\{1, 2\}, \{3\}\}, \ u_1 := x_1 + x_2 + 3x_3,$$
$$e_2 := (2, 0, 2), \ \mathcal{P}_2 := \{\{1, 3\}, \{2\}\}, \ u_2 := x_1 + 3x_2 + x_3,$$
$$e_3 := (0, 2, 2), \ \mathcal{P}_3 := \{\{2, 3\}, \{1\}\}, \ u_3 := 3x_1 + x_2 + x_3.$$

The net trades

$$z_1 = (-1, -1, 4),$$
$$z_2 = (-2, 3, -2),$$
$$z_3 = (3, -2, -2),$$

lead to the core allocation

$$x = \begin{pmatrix} 1 & 0 & 6 \\ 0 & 3 & 0 \\ 3 & 0 & 0 \end{pmatrix} \in C(E).$$

Obviously, one has that $x^n \in C(\mathbb{E}^n)$ for all $n \in \mathbb{N}$. Nevertheless, x is not a competitive quasi-equilibrium allocation. To see this, consider first the case where every agent has positive income, i.e. $e_i + z_i$ maximizes u_i. Note that $p = (p_1, p_2, p_3) \geq 0$, as the prices seen by an agent have to be nonnegative and $\{3\} \in \mathcal{P}_1, \{2\} \in \mathcal{P}_2, \{1\} \in \mathcal{P}_3$. Then one calculates from the budget equations $pz_i = 0$ (these have to hold if $e_i + z_i$ maximizes u_i), that the price vector would have to be $p = (p_1, p_1, \frac{1}{2}p_1)$. As $p \neq 0$ we must have $p > 0$. The budget set of agent 1 becomes $\{(x_1, x_2, x_3) \mid x - e_1 \in X_1$ and $px = pe_1 = 4p_1 > 0\}$. But then the unique maximizer of u_1 on this budget set is $(0, 0, 8) \neq x_1$. Hence, x cannot be a competitive quasi-equilibrium in this case, because agent 1 has positive income $4p_1$. On the other hand, if there is an agent with zero income, one can distinguish three cases.

1. $pe_1 = 0 \Rightarrow 2p_1 + p_2 + 2p_3 = 0 \overset{p \geq 0}{\Rightarrow} p = 0$, *a contradiction.*

2. $pe_2 = 0 \Rightarrow 2p_1 + 2p_3 = 0 \overset{p \geq 0}{\Rightarrow} p_1 = p_3 = 0 \overset{p \neq 0}{\Rightarrow} p_2 > 0 \Rightarrow pe_3 = 2p_2 > 0 \Rightarrow x_3 = (3, 0, 0)$
   maximizes u_3 on $B_3(p)$, a contradiction as $p_1 = 0$.

3. $pe_3 = 0 \Rightarrow 2p_2 + 2p_3 = 0 \overset{p \geq 0}{\Rightarrow} p_2 = p_3 = 0 \overset{p \neq 0}{\Rightarrow} p_1 > 0 \Rightarrow pe_2 = 2p_1 > 0 \Rightarrow x_2 = (0, 3, 0)$
   maximizes u_2 on $B_2(p)$, a contradiction as $p_2 = 0$. $\Diamond$

The next theorem points out that under trade boundedness of the communication system any equilibrium allocation x is also in the core. Trade boundedness here assures that all possible deviations of subcoalitions $S \subset N$ from x are also informational feasible in the grand coalition. They are therefore excluded from the budget sets of the agents in S not for informational infeasibility, but because they cannot be afforded.

Theorem 19 *Let $\mathbb{E}$ be an exchange economy with asymmetric information and trade bounded communication system. If x is a competitive equilibrium allocation, then $x \in C(\mathbb{E})$.*

Proof. Let (x, p) be a competitive equilibrium of $\mathbb{E}$. Suppose that $x \notin C(\mathbb{E})$. As $x - e$ is informational feasible for the grand coalition that can only be if $(u_i(x_i))_{i \in N} \in intV(S)$ for a coalition $\emptyset \neq S \subseteq N$, which means that there is an allocation y^S such that $u_i(y_i^S) > u_i(x_i)$ for all $i \in S$. As $(y_i^S - e_i)_{i \in S} \in \prod_{i \in S} X_i^S$ and $\sum_{i \in S}(y_i^S - e_i) = y^S(S) - e(S) = 0$, $(y_i^S - e_i)_{i \in S}$ is under trade boundedness also informational feasible in the grand coalition. Thus,

$$u_i(y_i^S) > u_i(x_i) \Longrightarrow y_i^S \notin B_i(p) \Longrightarrow py_i^S > pe_i$$

for all $i \in S$. Summing up, that leads to

$$\sum_{i \in S} py_i^S > \sum_{i \in S} pe_i$$
$$\Longleftrightarrow \quad py^S(S) > pe(S)$$
$$\Longrightarrow \quad y^S(S) \neq e(S)$$

a violation of physical feasibility. Hence, $x \in C(\mathbb{E})$. ∎

Again trade boundedness is a vital assumption. Otherwise economies are easily constructed, where $\{e\}$ is a competitive equilibrium, but the core is empty, due to the fact that trading is only possible in strict subcoalitions of N.

If x is a competitive equilibrium allocation in $\mathbf{E}$, then it is straightforward that x^n is a competitive equilibrium allocation in the replica economy $\mathbf{E}^n$. This shows that for competitive equilibrium the only if part of Theorem 17 holds in economies with a trade bounded communication system.

Corollary 20 *Let $\mathbf{E}$ be an exchange economy with asymmetric information and trade bounded communication system. If x is a competitive equilibrium*
   allocation then $x^n \in C(\mathbf{E}^n)$ for all $n \in \mathbf{N}$.

3.4   Equal Treatment

One way to derive an existence result for competitive allocations in the case of full information (or only one state of the world) requires to show that agents of the same type are treated 'equally' in the core. Then a compactness argument and a version of Theorem 17 are used to show that a subsequence of the mean allocations $\bar{x}_n := \frac{1}{n}\sum_{j=1}^{n} x_{.j}$, $x_n \in C(\mathbf{E}^n)$, converges to a competitive equilibrium. Equal treatment ensures that $(\bar{x}_n)^n \in C(\mathbf{E}^n)$, and thus $\bar{x}_n \in C(\mathbf{E})$. I want to use a similar approach in Section 3.5. However, when trying to carry over the proof of the equal treatment property a problem arises. The enforcement of equal treatment in all core allocations relies on the possibility to redistribute parts of the mean allocation of one owns type to all agents of the other types, making them better off. Under the measurability constraints on the net trades imposed by informational feasibility, this can no longer be guaranteed for arbitrary communication systems. In fact, the following simple example shows a situation, where unequal treatment is present in the core, exactly for the described reason.

Example 21   *Consider a private information economy with 3 agents $i \in N = \{1, 2, 3\}$, one commodity and three states $\Omega = \{1, 2, 3\}$. Let the initial endowments, the (private) information and the utility functions be*

$$
\begin{aligned}
e_1 &: \ = (1, 1, 0), \ \mathcal{P}_1 := \{\{1, 2\}, \{3\}\}, \ u_1(x) := x_1 + x_2 + 10x_3, \\
e_2 &: \ = (1, 0, 1), \ \mathcal{P}_2 := \{\{1, 3\}, \{2\}\}, \ u_2(x) := x_1 + 10x_2 + x_3, \\
e_3 &: \ = (1, 1, 1), \ \mathcal{P}_3 := \{\{2, 3\}, \{1\}\}, \ u_3(x) := 10x_1 + x_2 + x_3.
\end{aligned}
$$

The net trades

$$
\begin{aligned}
z_{(1,1)} &= z_{(1,2)} = (-1, -1, 2), \\
z_{(2,1)} &= (-1, \frac{2}{10}, -1), z_{(2,2)} = (-1, \frac{38}{10}, -1) \\
z_{(3,1)} &= z_{(3,2)} = (2, -1, -1),
\end{aligned}
$$

lead to the allocation

$$x = \begin{pmatrix} 0 & 0 & 2 \\ 0 & 0 & 2 \\ 0 & \frac{2}{10} & 0 \\ 0 & \frac{38}{10} & 0 \\ 3 & 0 & 0 \\ 3 & 0 & 0 \end{pmatrix}.$$

The agents of type 2 are not treated equally in this allocation. Nevertheless, this allocation is in $C(E^2)$. To see this, note that clearly x is informational and physically feasible. Suppose, contrary to $x \in C(E^2)$, that there is a coalition $S \subseteq N \times \{1,2\}, S \neq \emptyset$, and an allocation y for S such that $u_i(y_{(i,j)}) > u_i(x_{(i,j)})$ for all $(i,j) \in S$.

The first step is to show that $S \neq N \times \{1,2\}$. The allocation $\bar{x} := \frac{1}{2}(x_{i1} + x_{i2})_{i \in N}$ can easily be seen to be in $C(E)$, as any agent gets the complete endowment of the resource which he prefers most. Now, if $S = N \times \{1,2\}$, then $\bar{y} := \frac{1}{2}(y_{i1} + y_{i2})_{i \in N}$ is an allocation for N and the linearity of the utility functions causes $\bar{y}$ to be a deviation to $\bar{x}$ in the economy E, contradicting $\bar{x} \in C(E)$.

The next step is to see that $\mathrm{pr}^1(S) = N$. Otherwise only agents of the same type can trade with each other, which will not give them the possibility to deviate. So $S \neq N \times \{1,2\}$ and $\mathrm{pr}_1(S) = N$.

As $S \neq N \times \{1,2\}$ the initial endowment of the coalition S satisfies $e(S) \leq (5,4,4)$. Assume for the moment that there are two agents of type 3 in the coalition S. Then $u_3(y_{(3,1)}) + u_3(y_{(3,2)}) > 30 + 30 = 60$ and, as the utility function u_3 is linear and monotone, $u_3(y_{(3,1)}) + u_3(y_{(3,2)}) = u_3(y_{(3,1)} + y_{(3,2)}) \leq u_3(e(S)) \leq u_3(5,4,4) = 58$, a contradiction. Hence, without loss of generality $S \subseteq N \times \{1,2\} - \{(3,2)\}$. This implies $e(S) \leq (5,3,3)$. By similar arguments one can now continue to show that $S \subseteq N \times \{1,2\} - \{(1,2),(2,2),(3,2)\}$, which together with $\mathrm{pr}_1(S) = N$ implies that $S = \{(1,1),(2,1),(3,1)\}$.

The final step is now to show that there is no allocation $y = (y_{(1,1)}, y_{(2,1)}, y_{(3,1)})$ for the coalition $S = \{(1,1),(2,1),(3,1)\}$ which is a deviation, i.e. that satisfies $u_1(y_{(1,1)}) > 20$, $u_1(y_{(2,1)}) > 2$ and $u_1(y_{(3,1)}) > 30$. The problem of finding such an allocation y can be rewritten in the following way. The allocation y satisfies

$$y = \begin{pmatrix} 1 + z_{11} & 1 + z_{12} & z_{13} \\ 1 + z_{21} & z_{22} & 1 + z_{23} \\ 1 + z_{31} & 1 + z_{32} & 1 + z_{33} \end{pmatrix}.$$

Informational feasibility requires that $z_{11} = z_{12}$, $z_{21} = z_{23}$ and $z_{32} = z_{33}$. Physical feasibility requires that $z_{31} = -z_{11} - z_{21}$, $z_{32} = -z_{12} - z_{22}$ and $z_{13} = -z_{23} - z_{33}$. Combining this one obtains

$$y = \begin{pmatrix} 1 + z_{11} & 1 + z_{11} & z_{11} + z_{22} - z_{21} \\ 1 + z_{21} & z_{22} & 1 + z_{21} \\ 1 - z_{11} - z_{21} & 1 - z_{11} - z_{22} & 1 - z_{11} - z_{22} \end{pmatrix}.$$

So the rewritten problem is to find (z_{11}, z_{21}, z_{22}) such that $y = y(z_{11}, z_{21}, z_{22}) \geq 0$, $y(S) = e(S) = (3, 2, 2)$ and

$$
\begin{aligned}
\bar{u}_1 &: = u_1(y_{(1,1)}) = 2 + 12z_{11} - 10z_{21} + 10z_{22} > 20, \\
\bar{u}_2 &: = u_2(y_{(2,1)}) = 2 + 2z_{21} + 10z_{22} > 2, \\
\bar{u}_3 &: = u_3(y_{(3,1)}) = 12 - 12z_{11} - 10z_{21} - 2z_{22} > 30.
\end{aligned}
$$

The last three inequalities correspond to solving the linear equation system

$$
\begin{pmatrix} 12 & -10 & 10 \\ & 2 & 10 \\ -12 & -10 & -2 \end{pmatrix}
\begin{pmatrix} z_{11} \\ z_{21} \\ z_{22} \end{pmatrix} =
\begin{pmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \hat{u}_3 \end{pmatrix},
$$

subject to $\hat{u}_1 > 18, \hat{u}_2 > 0$ and $\hat{u}_3 > 18$. Now solving this system leads to

$$
\begin{aligned}
z_{11} &= \frac{4\hat{u}_1 - 5\hat{u}_2 - 5\hat{u}_3}{108} \\
z_{21} &= \frac{-5\hat{u}_1 + 4\hat{u}_2 - 5\hat{u}_3}{108}, \\
z_{22} &= \frac{\hat{u}_1 + 10\hat{u}_2 + \hat{u}_3}{108}.
\end{aligned}
$$

Physical feasibility implies that $1 + z_{21} \geq 0 \Leftrightarrow \hat{u}_2 \geq \frac{5\hat{u}_1 + 5\hat{u}_3 - 108}{4}$, and $\hat{u}_1 > 18, \hat{u}_3 > 18$ implies that $\frac{5\hat{u}_1 + 5\hat{u}_3 - 108}{4} > \frac{180 - 108}{4} = 18$, so $\hat{u}_2 > 18$. Hence, $y_{(2,1),2} = z_{22} = \frac{\hat{u}_1 + 10\hat{u}_2 + \hat{u}_3}{108} > \frac{216}{108} = 2 = e_2(S)$, a contradiction, since no agent can have more of a commodity than the complete endowment. So there is no deviation y for the coalition S. This concludes the proof that $x \in C(\mathbb{E}^2)$. ◇

Fortunately, what is really needed for the compactness argument is not equal treatment in all core allocations, but, that there is at least one core allocation $x \in C(\mathbb{E}^n)$, in which agents are treated equally. This can be guaranteed under the conditions of the following theorem.

Theorem 22 *Let* E *be an exchange economy with asymmetric information and trade bounded communication system. Assume that* $e_i \in \bigcap_{N \supseteq S \ni i} X_i^S$ *and* $u_i \in U_{qc}^{mo}$ *for all* $i \in N$. *Then there is an* $x \in C(\mathbb{E}^n)$, $n \in \mathbb{N}$, *in which agents of the same type receive the same commodity bundle.*

Instead of proving this theorem directly, we prove Lemma 23 for concave utilities, and explain how to obtain Theorem 22 from this lemma. Since this is a standard argument, it is not given in detail. The lemma shows, that in the special case of Theorem 22 where the utility functions are concave and normalized such that $u_i(0) = 0$, even a stronger result holds, namely that the mean allocation $\bar{x}^n$ of **any** core allocation x is also in the core.

Lemma 23 *Let E be an exchange economy with asymmetric information and trade bounded communication system. Assume that $e_i \in \bigcap_{S \ni i} X_i^S$ and $u_i \in U_{co,0}^{mo}$ for all $i \in N$. Then $x \in C(E^n)$, $n \in \mathbf{N}$, implies that $\overline{x}^n$ is also in $C(E^n)$.*

The proof of the lemma relies on some other observations.

Lemma 24 *Let E be an exchange economy with asymmetric information and trade bounded communication system. Assume that $u_i \in U_{qc}$ for all $i \in N$. For any assignment of commodities $x : \Omega \to \mathbf{R}_+^{n \times k}$ to the agents $1, \ldots, n$, $n \geq 2$, of a given type $i \in N$, and any $S \subseteq \{1, \ldots, n\}$, $|S| \geq 2$, there is an $l \in S$ such that*

$$u_i\left(\frac{1}{|S|}\sum_{j \in S} x_j\right) \geq u_i\left(\frac{1}{|S|-1}\sum_{j \in S, j \neq l} x_j\right).$$

Proof. Suppose to the contrary that

$$u_i\left(\frac{1}{|S|-1}\sum_{j \in S, j \neq l} x_j\right) > u_i\left(\frac{1}{|S|}\sum_{j \in S} x_j\right)$$

for all $l \in S$. Then quasi-concavity yields

$$u_i\left(\frac{1}{|S|}\sum_{l \in S}\left(\frac{1}{|S|-1}\sum_{j \in S, j \neq l} x_j\right)\right) \geq \min_{l \in S} u_i\left(\frac{1}{|S|-1}\sum_{j \in S, j \neq l} x_j\right)$$

$$> u_i\left(\frac{1}{|S|}\sum_{j \in S} x_j\right).$$

But the operand on the left hand side, $\frac{1}{|S|}\sum_{l \in S}\frac{1}{|S|-1}\sum_{j \in S, j \neq l} x_j$, actually equals the operand of the right hand side, $\frac{1}{|S|}\sum_{j \in S} x_j$, a contradiction. ∎

This lemma allows to order $(x_1, \ldots, x_n)$ in a specific way.

Lemma 25 *Let E be an exchange economy with asymmetric information and trade bounded communication system. Assume that $u_i \in U_{qc}$ for all $i \in N$. For any assignment of commodities $x : \Omega \to \mathbf{R}_+^{n \times k}$ to the agents $1, \ldots, n$, $n \geq 2$, of a given type $i \in N$, there is a permutation π of $(1, \ldots, n)$ such that for $x^\pi := (x_{\pi(1)}, \ldots, x_{\pi(n)})$ it holds that*

$$u_i\left(\frac{1}{n-m+1}\sum_{j=m}^{n} x_j^\pi\right) \geq u_i\left(\frac{1}{n-m}\sum_{j=m+1}^{n} x_j^\pi\right)$$

for all $m = 1, \ldots, n-1$.

Proof. π can be constructed inductively. Apply Lemma 24 to

$$S = \{1, \ldots, n\}$$

and let $\pi(1) = l$. When $\pi(1), \ldots, \pi(j)$ are constructed for some $2 \leq j < n$, apply Lemma 24 to

$$S = \{1, \ldots, n\} \setminus \{\pi(1), \ldots, \pi(j)\},$$

and let $\pi(j+1) = l$. When finally $j = n$, let $\pi(n)$ be the last remaining element in

$$S = \{1, \ldots, n\} \setminus \{\pi(1), \ldots, \pi(n-1)\}.$$

Note that the $\pi(j)$, $1 \leq j < n$, obtained in this way are distinct so that the process really ends with $|\{1, \ldots, n\} \setminus \{\pi(1), \ldots, \pi(n-1)\}| = 1$ and π is a permutation of $(1, \ldots, n)$. The desired property holds by the construction via Lemma 24. ∎

Now the proof of Lemma 23 can be given.

Proof. Assume to the contrary that $\bar{x}^n$ is not in $C(E^n)$. So there is a subcoalition S of the grand coalition $N \times \{1, \ldots, n\}$ and an allocation y^S for that coalition, such that

$$u_i(y^S_{(i,j)}) > u_i(\bar{x}^n_{(i,j)}) \text{ for all } (i,j) \text{ in } S.$$

Using concavity one obtains therefore for

$$\bar{y}^S_{(i,j)} := \frac{1}{|pr_i(S)|} \sum_{j \in pr_i(S)} y^S_{(i,j)},$$

that

$$u_i(\bar{y}^S_{(i,j)}) \geq \frac{1}{|pr_i(S)|} \sum_{j \in pr_i(S)} u_i(y^S_{(i,j)}) > u_i(\bar{x}^n_{(i,j)}) \text{ for all } (i,j) \text{ in } S.$$

It is easily seen, that $\bar{y}^S$ is also physically and informational feasible for the coalition S. As $\bar{x}^n_{(i,j)} = \bar{x}^n_{(i,k)}$ for all $k \neq j$ one can without loss of generality assume that $pr_i(S) = \{l_i, \ldots, n\}$, $1 \leq l_i \leq n$, $i \in N$. Furthermore, $x \in C(E^n)$ implies that for any collection of permutations $(\pi_i)_{i \in N}$ of $(1, \ldots, n)$ $(x_i^{\pi_i})_{i \in N} := (x_{(i, \pi_i(j))})_{i,j}$ is also in $C(E^n)$. Additionally permutations do not change the mean $\bar{x}^n_{(i,j)}$, so it can be assumed by Lemma 25 that

$$
\begin{aligned}
u_i(\bar{x}^n_{(i,j)}) &= u_i\left(\frac{1}{n} \sum_{j=1}^{n} x_{(i,j)}\right) \\
&\geq u_i\left(\frac{1}{n-1} \sum_{j=2}^{n} x_{(i,j)}\right) \\
&\geq \cdots \\
&\geq u_i\left(\frac{1}{n-l_i+1} \sum_{j=l_i}^{n} x_{(i,j)}\right) \\
&= u_i\left(\frac{1}{|pr_i(S)|} \sum_{j \in pr_i(S)} x_{(i,j)}\right).
\end{aligned}
$$

Hence, altogether this yields

$$u_i(\bar{y}^S_{(i,j)}) > u_i\left(\frac{1}{|pr_i(S)|} \sum_{j \in pr_i(S)} x_{(i,j)}\right) \geq \frac{1}{|pr_i(S)|} \sum_{j \in pr_i(S)} u_i(x_{(i,j)}). \tag{3.1}$$

Concavity and $u_i(0) = 0$ imply that

$$u_i(\lambda x) = u_i(\lambda x + (1-\lambda)0) \geq \lambda u_i(x) + (1-\lambda)u_i(0) = \lambda u_i(x) \qquad (3.2)$$

for all $\lambda \in [0, 1]$. Equation 3.1 is equivalent to

$$\sum_{j \in pr_i(S)} u_i(\overline{y}^S_{(i,j)}) > \sum_{j \in pr_i(S)} u_i(x_{(i,j)}) \geq u_i(x_{(i,\overline{j})}) \text{ for any } \overline{j} \in pr_i(S),$$

as $u_i \geq 0$ on $\mathbb{R}^{k \times \Omega}_+$ ($u_i \in U^{mo}_{co,0}$). Note that $u_i(\overline{y}^S_{(i,j)}) > 0$, so one can define

$$\lambda_{(i,j)} := \frac{u_i(x_{(i,j)})}{u_i(\overline{y}^S_{(i,j)})},$$

and using equation 3.2 one has

$$u_i(\lambda_{(i,j)}\overline{y}^S_{(i,j)}) \geq \lambda_{(i,j)}u_i(\overline{y}^S_{(i,j)}) = \frac{u_i(x_{(i,j)})}{u_i(\overline{y}^S_{(i,j)})}u_i(\overline{y}^S_{(i,j)}) = u_i(x_{(i,j)})$$

for all $i \in pr^1(S)$, $j \in pr_i(S)$.

Furthermore,

$$\sum_{j \in pr_i(S)} \lambda_{(i,j)}\overline{y}^S_{(i,j)} = \left(\sum_{j \in pr_i(S)} \lambda_{(i,j)}\right)\overline{y}^S_{(i,j)}$$

$$= \left(\sum_{j \in pr_i(S)} \frac{u_i(x_{(i,j)})}{u_i(\overline{y}^S_{(i,j)})}\right)\overline{y}^S_{(i,j)}$$

$$= |pr_i(S)|\underbrace{\left(\frac{\sum_{j \in pr_i(S)} u_i(x_{(i,j)})}{\sum_{j \in pr_i(S)} u_i(\overline{y}^S_{(i,j)})}\right)}_{<1 \text{ by equation 3.1}}\overline{y}^S_{(i,j)}$$

$$< |pr_i(S)|\overline{y}^S_{(i,j)} = \sum_{j \in pr_i(S)} \overline{y}^S_{(i,j)}.$$

Hence, letting for all $i \in pr^1(S)$

$$y^r_i := \sum_{j \in pr_i(S)} \overline{y}^S_{(i,j)} - \sum_{j \in pr_i(S)} \lambda_{(i,j)}\overline{y}^S_{(i,j)} = \sum_{j \in pr_i(S)} (1 - \lambda_{(i,j)})\overline{y}^S_{(i,j)},$$

one has that $y^r_i > 0$ and $y^r_i \in X^S_i$. Let

$$\widetilde{y}^S_{(i,j)} := \lambda_{(i,j)}\overline{y}^S_{(i,j)} + \frac{y^r_i}{|pr_i(S)|}.$$

The proof is finished, if it can be shown that $\widetilde{y}^S$ is an allocation for S, because then, as

$$u_i(\widetilde{y}^S_{(i,j)}) > u_i(x_{(i,j)}) \text{ for all } (i,j) \text{ in } S,$$

$\tilde{y}^S$ is a deviation to x, contradicting $x \in C(E^n)$. $\tilde{y}^S$ is an allocation for S, if it is informational and physically feasible. Informational feasibility and $\tilde{y}^S \geq 0$ follows from

$$
\begin{aligned}
\tilde{y}_{(i,j)}^S \;=\;& \lambda_{(i,j)}\bar{y}_{(i,j)}^S + \frac{y_i^r}{|pr_i(S)|}. \\[4pt]
=\;& \lambda_{(i,j)}\bar{y}_{(i,j)}^S + \frac{1}{|pr_i(S)|} \sum_{j \in pr_i(S)} (1 - \lambda_{(i,j)})(e_{(i,j)} + \bar{y}_{(i,j)}^S - e_{(i,j)}) \\[4pt]
=\;& e_{(i,j)} + \underbrace{\lambda_{(i,j)}\bar{y}_{(i,j)}^S}_{\in X_i^S} + \frac{1}{|pr_i(S)|} \sum_{j \in pr_i(S)} \underbrace{(1 - \lambda_{(i,j)})(\bar{y}_{(i,j)}^S - e_{(i,j)})}_{\in X_i^S} \\[4pt]
& \underbrace{- \frac{e_{(i,j)}}{|pr_i(S)|} \underbrace{\sum_{j \in pr_i(S)} \lambda_{(i,j)}}_{< |pr_i(S)|}}_{\in X_i^S} \\[4pt]
>\;& e_{(i,j)} + \lambda_{(i,j)}\bar{y}_{(i,j)}^S + \frac{1}{|pr_i(S)|} \sum_{j \in pr_i(S)} \underbrace{(1 - \lambda_{(i,j)})(\bar{y}_{(i,j)}^S - e_{(i,j)})}_{\geq 0} - e_{(i,j)} \\[4pt]
\geq\;& 0.
\end{aligned}
$$

The remaining part of physical feasibility follows from

$$
\begin{aligned}
\sum_{(i,j) \in S} \tilde{y}_{(i,j)}^S \;=\;& \sum_{i \in pr^1(S)} \sum_{j \in pr_i(S)} \left(\lambda_{(i,j)}\bar{y}_{(i,j)}^S + \frac{y_i^r}{|pr_i(S)|} \right) \\[4pt]
=\;& \sum_{i \in pr^1(S)} \sum_{j \in pr_i(S)} \left(\lambda_{(i,j)}\bar{y}_{(i,j)}^S + \frac{1}{|pr_i(S)|} \sum_{j \in pr_i(S)} (1 - \lambda_{(i,j)})\bar{y}_{(i,j)}^S \right) \\[4pt]
=\;& \sum_{i \in pr^1(S)} \sum_{j \in pr_i(S)} \left(\lambda_{(i,j)}\bar{y}_{(i,j)}^S + (1 - \lambda_{(i,j)})\bar{y}_{(i,j)}^S \right) \\[4pt]
=\;& \sum_{(i,j) \in S} \bar{y}_{(i,j)}^S = \sum_{(i,j) \in S} y_{(i,j)}^S = e(S).
\end{aligned}
$$

∎

To prove Theorem 22 from Lemma 23, use Hildenbrand (1988), p.58, Proposition A.2, to approximate $u \in U_{qc}^{mo}$ by a sequence of $(u_j)_{j \in N}$ such that $u_j \in U_{co,0}^{mo}$ for all $j \in N$. Then use upper hemicontinuity of the core, viewed as a compact-valued correspondence from $(N, \Omega, e, (P_i^S)_{i \in S, S \subseteq N}) \times (U_{qc}^{mo})^N$ to $\mathbb{R}^{N \times k \times \Omega}$, i.e. when the utility functions are variables with domain U_{qc}^{mo}. Any sequence of elements from the cores of the approximating economies yields a sequence satisfying equal treatment by Lemma 23, and thus by upper hemicontinuity a core element of the quasi-concave economy.

3.5 Existence of competitive allocations

As a consequence of Theorem 12 one obtains in combination with Theorem 17 an existence result for the quasi-equilibrium.

Theorem 26 *Let E be an exchange economy with asymmetric information and trade bounded communication system. Assume that $e_i \in \bigcap_{S \ni i} X_i^S$ and $u_i \in U_{qc}^{me} \cap C(\mathbb{R}_+^{k \times \Omega})$ for all $i \in N$. Then this economy has a competitive quasi-equilibrium (p, x). Moreover the price system p can be chosen $\sigma(\bigcup_{i \in N} \mathcal{P}_i^N)$-measurable.*

Proof. By Theorem 12 $C(\mathbf{E}^n) \neq \emptyset$ for all $n \in N$. By Theorem 17 it suffices to show that the set of all $x \in \mathbb{R}^{N \times k \times \Omega}$ such that $x^n \in C(\mathbf{E}^n)$ for all $n \in N$ is not empty. To see this, let $x_n \in C(\mathbf{E}^n)$ be a sequence of core elements from the replica economies $\mathbf{E}^n$, $n \in N$. By Theorem 22 it can be assumed that agents of the same type are treated equally, i.e. $x_n = (y_n)^n$ for all $n \in N$ and allocations y_n in the economy $\mathbf{E}$. Furthermore, $y_n \in C(\mathbf{E})$ for all $n \in N$, because any deviation to y_n in the economy $\mathbf{E}$ would also be a deviation to $(y_n)^n$ in the economy $\mathbf{E}^n$. As $C(\mathbf{E})$ is compact, there is a convergent subsequence $(y_{n_k})_k$, $n_k \to \infty$, of $(y_n)_n$. Let y be the limit of this subsequence. If it is shown that $y^n \in C(\mathbf{E}^n)$ for all $n \in N$ the proof is finished. Assume to the contrary that y^n is not in $C(\mathbf{E}^n)$ for some $n \in N$.

So there is coalition $S \subseteq N \times \{1, \dots, n\}$ and an allocation x^S for that coalition such that

$$u_{(i,j)}(x_{(i,j)}^S) > u_{(i,j)}(y_{(i,j)}^n) \text{ for all } (i,j) \in S.$$

As $|S| < \infty$, there is an $\varepsilon > 0$ such that

$$u_{(i,j)}(x_{(i,j)}^S) - u_{(i,j)}(y_{(i,j)}^n) > \varepsilon \text{ for all } (i,j) \in S.$$

On the other hand, using continuity of utilities and $y_{n_k} \to y$, there is a k_0, such that for all $k \geq k_0$,

$$\varepsilon \geq u_i(y_i) - u_i((y_i)_{n_k}) \geq -\varepsilon.$$

But then, as $u_{(i,j)}(y_{(i,j)}^n) = u_i(y_i)$,

$$
\begin{aligned}
&u_{(i,j)}(x_{(i,j)}^S) - u_{(i,j)}((y_{(i,j)})_{n_k}^{n_k}) \\
={} &u_{(i,j)}(x_{(i,j)}^S) \underbrace{- u_{(i,j)}(y_{(i,j)}^n)}_{=0} + u_i(y_i) - u_i((y_i)_{n_k}) \\
>{} &\varepsilon + (-\varepsilon) \geq 0 \text{ for all } (i,j) \in S.
\end{aligned}
$$

Hence, x^S is also an allocation for S in the economies $\mathbf{E}^{n_k}$, $k \geq k_0$, $n_k \geq n$, such that

$$u_{(i,j)}(x_{(i,j)}^S) > u_{(i,j)}((y_{(i,j)})_{n_k}^{n_k}) \text{ for all } (i,j) \in S.$$

However, this contradicts $(y_{n_k})^{n_k} \in C(\mathbf{E}^{n_k})$. ∎

Next situations are identified, where $\inf pX_i^{N, \geq 0} < pe_i$ is true for all $i \in N$, so that Theorem 26 turns into an existence result for the competitive equilibrium. As $\inf pX_i^{N, \geq 0}$ must be equal to 0 for strictly increasing utilities, it is the same to identify when every agent has positive income.

Let the *information matrix* $\mathbb{I}(S)$ of a coalition S be the matrix in $\{0,1\}^{\Omega \times \cup_{i \in S} \mathcal{P}_i}$, where the entry $\mathbb{I}(S)_{(\omega, A)}$ is equal to $\mathbb{I}_A(\omega)$, i.e.

$$\mathbb{I}(S) := (\mathbb{I}_A(\omega)_{\omega \in \Omega})_{A \in \cup_{i \in S} \mathcal{P}_i}.$$

This matrix maps

$$X_+ := \{x \in \mathbb{R}^{k \times \Omega} \mid x \text{ is } \sigma(\cup_{i \in N} \mathcal{P}_i)\text{-measurable and } x_A \geq 0 \text{ for all } A \in \cup_{i \in N} \mathcal{P}_i\}$$

to $\mathbb{R}^{k \times \cup_{i \in N} \mathcal{P}_i}$ in the following way. Let

$$x \mathbb{I}(S) := (x_j \mathbb{I}(S)_A)_{1 \leq j \leq k, A \in \cup_{i \in N} \mathcal{P}_i}.$$

The set of all elements in X_+ mapped to the zero matrix in $\mathbb{R}^{k \times \cup_{i \in N} \mathcal{P}_i}$ is then

$$K(\mathbb{I}(S)) := \{x \in X_+ \mid x\mathbb{I}(S) = 0\}.$$

Lemma 27 *Let* E *be an exchange economy with asymmetric information. Assume that* $e_i \in X_i^N$, $e_i \neq 0$, $\sum_{i \in N} e_i > 0$, *and* $u_i \in U^{mo}$ *for all* $i \in N$. *Each of (i) to (iv) is a sufficient condition for every agent to have positive income in a quasi-equilibrium* (x,p) *of this economy, where* p *is* $\sigma(\cup_{i \in N} \mathcal{P}_i^N)$*-measurable.*

(i) *The communication system is symmetric for the grand coalition, i.e.* $\mathcal{P}_i^N \equiv \mathcal{P}^N$ *for all* $i \in N$.

(ii) $pe_i > 0$ *for some agent* $i \in N$, *and all other agents have positive initial endowments.*

(iii) $K(\mathbb{I}(S)) = \{0\}$ *for some coalition* S, *and all agents have positive initial endowments.*

(iv) $K(\mathbb{I}(S)) = \{0\}$ *for some coalition* S, *all agents in* S *have positive initial endowments, and this coalition knows collectively more than the agents in* $N \setminus S$, *i.e.*

$$\bigcup_{i \in S} \mathcal{P}_i \supseteq \bigcup_{i \in N \setminus S} \mathcal{P}_i.$$

Proof. Let (x,p) be a quasi-equilibrium of the economy E.

(i): In such a communication system the set of all $\sigma(\cup_{i \in N} \mathcal{P}_i)$-measurable price systems $p \neq 0$ such that $p(A) \geq 0$ for all $A \in \mathcal{P}_i$ is equal to the set $X_i^{N, \geq 0} \setminus \{0\}$ for any $i \in N$. So, as $\sum_{i \in N} e_i > 0$, at least one agent i has positive income. But then $p(A) > 0$ for all $A \in \mathcal{P}_i^N = \mathcal{P}_i$, $i \in N$, so every agent has positive income.

(ii): As $pe_i > 0$ leads to $p(A) > 0$ for all $A \in \mathcal{P}_i$, it can be concluded that $p(\Omega) > 0$. But then, as e_j is $\sigma(\mathcal{P}_j^N)$-measurable and $p(A) \geq 0$ for any $A \in \mathcal{P}_j^N$, $j \neq i$,

$$pe_j = \sum_{A \in \mathcal{P}_j^N} p(A) e_j(A)$$

$$\geq \sum_{A \in \mathcal{P}_j^N} p(A) \min_{A \in \mathcal{P}_j^N} e_j(A)$$

$$= \left(\sum_{A \in \mathcal{P}_j^N} p(A) \right) \min_{A \in \mathcal{P}_j^N} e_j(A)$$

$$= \underbrace{p(\Omega)}_{>0} \underbrace{\min_{A \in \mathcal{P}_j^N} e_j(A)}_{>0}$$

$$> 0,$$

hence any agent has positive income.

(iii): I show that $K(\mathbb{I}(S)) = \{0\}$ implies that one agent in S has positive income, then (ii) can be used. So suppose to the contrary, that

$$pe_i = \sum_{A \in \mathcal{P}_i} p(A) e_i(A) = 0 \text{ for all } i \in S.$$

Then, as $p(A) \geq 0$ and $e_i(A) \geq 0$ for all $A \in \cup_{i \in N} \mathcal{P}_i$, one obtains that

$$p(A) e_i(A) = 0 \text{ for all } A \in \cup_{i \in S} \mathcal{P}_i.$$

As $e_i(A) > 0$ for all $A \in \cup_{i \in S} \mathcal{P}_i$ this leads to

$$p(A) = 0 \text{ for all } A \in \cup_{i \in S} \mathcal{P}_i,$$

or, because $p(A) \geq 0$ for all $A \in \cup_{i \in N} \mathcal{P}_i$ and p is $\sigma(\cup_{i \in N} \mathcal{P}_i)$-measurable,

$$p \in K(\mathbb{I}(S)) = \{0\}.$$

This contradicts $p \neq 0$.

(iv): As in (iii), one agent in S must have positive income. But then $p(\Omega) > 0$, so all agents in S have positive income, showing that $p(A) > 0$ for all $A \in \cup_{i \in S} \mathcal{P}_i$. As $\cup_{i \in S} \mathcal{P}_i \supseteq \cup_{i \in N \setminus S} \mathcal{P}_i$ and $e_i \geq 0$, $e_i \neq 0$ for all $i \in N \setminus S$, this means that the agents in $N \setminus S$ also have positive income. ∎

Note that the situations described in (i), (iii) and (iv) are independent of the given quasi-equilibrium (x, p). So the following corollary to Theorem 26 can be stated.

Corollary 28 *Let* **E** *be an exchange economy with asymmetric information and trade bounded communication system. Assume that $e_i \in \cap_{S \ni i} X_i^S$ and $u_i \in U_{qc}^{mo} \cap C(\mathbb{R}_+^{k \times \Omega})$ for all $i \in N$. Let the assumptions and one of the situations (i), (iii) or (iv) given in Lemma 27 hold. Then this economy has a competitive equilibrium, and any competitive quasi-equilibrium is already a competitive equilibrium.*

As stated after Definition 14, this corollary is able to cover existence of a Radner equilibrium without free disposal. Other existence results are contained in Einy et al. (2001), for the case of a non-atomic measure space of agents, and Radner (1968, 1982), for the case of a finite number of agents. All of these papers assume some degree of disposability, in Einy et al. (2001) and Radner (1982) free disposal, and in Radner (1968) production sets have a nonempty relative interior, which translates into disposability of at least part of the endowments in a pure exchange economy. It is known that if one wants to dispense with free disposal one has to allow for negative prices. An example of an economy without free disposal such that equilibrium prices are negative can be found in Liu (1992). The prices as defined here are consequently also allowed to be negative. Nevertheless, in equilibrium all prices seen by agents will have to be positive, so that nobody can exploit the existence of negative prices. In contrast to the machinery of proof used in this chapter the mentioned existence results are not obtained by appealing to a limit theorem for the core. Moreover, Radner has to assume concavity of state-dependent utility in order to achieve concavity of expected utility. This corresponds to concavity of the utility functions $u_i : \mathbb{R}^{k \times \Omega}_+ \to \mathbb{R}$ used here. No such assumptions are necessary in Einy et al. (2001), as they are dealing with a non-atomic measure space of agents.

Part II

Partially defined games

As stated after Definition 14, this corollary is able to cover existence of a Radner equilibrium without free disposal. Other existence results are contained in Einy et al. (2001), for the case of a non-atomic measure space of agents, and Radner (1968, 1982), for the case of a finite number of agents. All of these papers assume some degree of disposability, in Einy et al. (2001) and Radner (1982) free disposal, and in Radner (1968) production sets have a nonempty relative interior, which translates into disposability of at least part of the endowments in a pure exchange economy. It is known that if one wants to dispense with free disposal one has to allow for negative prices. An example of an economy without free disposal such that equilibrium prices are negative can be found in Liu (1992). The prices as defined here are consequently also allowed to be negative. Nevertheless, in equilibrium all prices seen by agents will have to be positive, so that nobody can exploit the existence of negative prices. In contrast to the machinery of proof used in this chapter the mentioned existence results are not obtained by appealing to a limit theorem for the core. Moreover, Radner has to assume ... state-dependent utility in order to achieve concavity of expected utility. This corresponds to concavity of the utility function ... No such assumptions are necessary in Einy et al. (2001), as they are ...

Chapter 4

The expected Shapley value

A method to allocate the benefits to the players of a cooperative game is the Shapley value. Its computation demands the knowledge of all coalition worths with certainty. This chapter introduces the expected Shapley value, an extension of the Shapley value to games where not all the worths are known with certainty. The expected Shapley value is characterized with adapted versions of Young's (1985) and Shapley's (1953) properties. It is shown how representation with unanimity games, dividend and potential generalize when not all worths are known with certainty. We relate the expected Shapley value to the reduced and the normalized Shapley value introduced by Willson (1993) and Housman (2001) for games where some coalition worths are not known.[1]

4.1   Introduction

The Shapley value (Shapley, 1953) is a method to allocate benefits to the players of a cooperative game. However, the Shapley value requires a large amount of information, since it takes into account the worth of any coalition. As the number of coalitions grows exponentially with the number of players, one might want to consider cooperative games where only a part of these worths are known. Such games are called partially defined games, and the coalitions whose worths are known in that game are collected in a coalition list, which is a subset of the set of all coalitions.

Willson (1993) and Housman (2001) characterized extensions of the Shapley value to partially defined games with symmetric coalition lists: b-weighted Shapley value, reduced Shapley value and normalized Shapley value. These extensions are based on extensions of axioms of efficiency, anonymity, linearity and dummy-player. The original axioms characterize the Shapley value for cooperative games. Housman (2001), however, obtains several impossibility results, in terms of coalition lists admitting such extensions of the Shapley value. These impossibilities suggest that the axioms used are too strong, which is made

---

[1]The author thanks particularly Hans Peters and Ton Storcken for helpful suggestions. The results in this chapter were first formulated in Maus (2003).

more explicit by an example in Section 4.2. In particular linearity is too strong, as stated
already in the conclusion of Housman (2001). The reason is that any extension of the sum
of two games can arise in multiple ways from extensions of the two original games, but for
some extensions of the sum there are more such ways than for others.

The path we take in this chapter differs somewhat from that of Willson (1993) and
Housman (2001). We use their concept of a partially defined cooperative game with coali-
tion list. Additionally we assume that players have a belief on the underlying cooperative
games, where all worths are known. This belief could arise from players who believe that
they are playing a game of a certain type, for example a superadditive game, and consider
any underlying superadditive extension *a priori* equally likely. These beliefs in turn allow
us to characterize the expected Shapley value. This value of a partially defined cooperative
game can be argued to arise in the following way. First, given the belief of the players,
compute the Shapley value of any of the games that could be the underlying cooperative
game. Integrate these values with respect to the probability distribution given by the belief
of the players. The resulting value is the expected Shapley value. The expected Shapley
value encompasses the reduced and normalized Shapley value as special cases, i.e. for a
certain type of belief. The approach used here also allows for any coalition list, possibly
asymmetric, containing the empty coalition. The price we have to pay for the expected
Shapley value is, that, compared to the reduced Shapley value and normalized Shapley
value, more information is required since beliefs about the underlying games are needed.
But this is still a weaker requirement than knowing all worths with certainty. A common
assumption in practise might be to take uniform distributions over games with standard
properties, such as convexity, superadditivity or monotonicity.

In the next section the relevant definitions are given and an example of a nonlinear
solution is discussed. Section 4.3 contains a characterization of the expected Shapley value
with a standard set of properties *à la* Young (1985), of which actually only monotonicity
has to be adapted to take into account beliefs. After that Section 4.4 explores a weaker
form of linearity, which is used in another characterization of the expected Shapley value,
this time with a standard set of properties *à la* Shapley (1953). Then Section 4.5 explains
how games with beliefs can be represented by unanimity games, and how the concepts of
dividend and potential extend. In Section 4.6 we show how to obtain the reduced and the
normalized Shapley value as expected Shapley values. Finally, Section 4.7 concludes the
chapter.

4.2 Partially defined cooperative games

Throughout, when vectors are compared, this is done coordinatewise.

A set of *players* N is a finite nonempty subset of the natural numbers $\mathbb{N}$. A subset S of
N is called a *coalition*, and we write $|S|$ for the number of players in the coalitions. Next
we introduce the class of partially defined cooperative games. Let $\mathcal{M} \subseteq 2^N, \emptyset \in \mathcal{M}$ be a

subset of coalitions, then $\mathcal{M}$ is called a *coalition list in* N. A pair (N, v) consisting of a function $v : \mathcal{M} \to \mathbb{R}$ such that $v(\emptyset) = 0$ and a player set N is called a *partially defined cooperative game* with coalition list $\mathcal{M}$. We will also call it an $\mathcal{M}$-game. We call $v(S)$ the *worth* of coalition S if S is in $\mathcal{M}$. If S is not in $\mathcal{M}$ then the interpretation is that the worth of this coalition is

not known. It might be that worths are costly to determine, for example. A coalition list $\mathcal{M}$ is *symmetric*, if for any $S \in \mathcal{M}$, any coalition of the same cardinality is also in $\mathcal{M}$. It is easily seen that if $\mathcal{M}$ is symmetric, then there is a set of *known coalition sizes* $K \subseteq \{0, \ldots, |N|\}$ such that

$$\mathcal{M} = \mathcal{M}(K) := \{S \subseteq N \mid |S| = k \text{ for some } k \in K\}.$$

If there is no confusion the $\mathcal{M}$-game (N, v) is referred to by v. Note that it is not necessary to define a partially defined cooperative game as a triple $(N, \mathcal{M}, v)$, since $\mathcal{M} = dom(v)$ is given by the domain of v.[2]

The set of all partially defined games with player set N is denoted by PG^N. So

$$PG^N = \{v : \mathcal{M} \to \mathbb{R} \mid \mathcal{M} \text{ is a coalition list in player set } N, v(\emptyset) = 0\}.$$

A *cooperative game* with player set N is a partially defined cooperative game (N, w), where w has domain $\mathcal{M} = 2^N$. Since then all coalition worths are known we will also refer to cooperative games as *fully defined games*. The set of all cooperative games with player set N is denoted by G^N. A *value solution* on G^N is a function $f : G^N \to \mathbb{R}^N$. The number $f_i(w)$ is interpreted as the share of player i in the game $w \in G^N$. The Shapley value $Sh : G^N \to \mathbb{R}^N$ for a cooperative game $w \in G^N$ is defined by

$$Sh_i(w) := \sum_{i \notin S, S \subseteq N} \frac{|S|!(|N| - 1 - |S|)!}{|N|!} \left(w(S \cup \{i\}) - w(S)\right),$$

for all $i \in N$. In this chapter we aim to extend to Shapley value on G^N to partially defined cooperative games. The set G^N of fully defined games with player set N is a subset of PG^N. The set of all fully defined games is given by

$$G := \bigcup \{G^N \mid \emptyset \neq N \subseteq \mathbb{N}, |N| < \infty\}.$$

Likewise, the set of all partially defined games is given by

$$PG := \bigcup \{PG^N \mid \emptyset \neq N \subseteq \mathbb{N}, |N| < \infty\}.$$

Again, G is a subset of PG.

Let (N, v) be a partially defined cooperative game with coalition list $\mathcal{M}$. Any fully defined game $w \in G^N$ satisfying $w|_{\mathcal{M}} = v$ is called an *extension* of v.

---

[2]Any function is also a relation, so $v \subseteq 2^N \times \mathbb{R}$ and the domain of v is given by the projection on the first coordinate of v.

Some important subspaces of the set of all fully defined games are the *monotonic games*

$$MO^N := \{w \in G^N \mid w(S) \le w(T) \text{ for all } S, T \in 2^N \text{ with } S \subseteq T\},$$

the *superadditive games*

$$SA^N := \{w \in G^N \mid w(S) + w(T) \le w(S \cup T) \text{ for all } S, T \in 2^N \text{ with } S \cap T = \emptyset\}$$

and the *convex games*

$$CO^N := \{w \in G^N \mid w(S) + w(T) \le w(S \cup T) + w(S \cap T) \text{ for all } S, T \in 2^N\}.$$

We proceed with an example, where players believe that each game from a certain set of extensions to fully defined games has equal probability to be the underlying game. This gives rise to a nonlinear solution to the allocation problem. Thus the example shows that linearity cannot be adapted in a straightforward fashion. The reason is that when two games are added, any extension of the sum of these games can be expressed as a sum of extensions of the two games added, but some extensions are 'more likely' than others to arise in this way.

Example 29 *There are three players $N = \{1, 2, 3\}$. The coalition list is*

$$\mathcal{M} = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, N\}.$$

So only the worths of the coalitions $\{1, 3\}$ and $\{2, 3\}$ are unknown. Let v be the $\mathcal{M}$-game satisfying

$$
\begin{aligned}
v(N) &= 1, \\
v(\{i\}) &= 0 \text{ for all } i \in N, \\
v(\{1, 2\}) &= 0.
\end{aligned}
$$

Suppose that players believe that they are playing a convex game. The constraints for extensions w derived from the convexity conditions are

$$w(\{1, 3\}) + w(\{2, 3\}) \le v(N) = 1,$$

$$w(\{1, 3\}) \ge 0 \text{ and } w(\{2, 3\}) \ge 0.$$

Hence, the set of convex extensions of v can be identified with the set

$$W := \{(x, y) \mid x, y \ge 0 \text{ and } x + y \le 1\}.$$

Let $\hat{v}$ be an $\mathcal{M}$-game which equals v except for $\hat{v}(\{1, 2\}) = \frac{1}{2}$. Then for any convex extension $\hat{w}$ we have that $0 \le \hat{w}(\{1, 3\}) \le \frac{1}{2}$ and $0 \le \hat{w}(\{2, 3\}) \le \frac{1}{2}$. This set of extensions can be identified with the set

$$\widehat{W} := [0, \frac{1}{2}]^2.$$

The players consider each game equally likely so they might agree to compute the expected games E_v *and* $E_{\hat{v}}$ *under the uniform distribution on* W *and* $\widehat{W}$, *respectively, and distribute according to the Shapley value of these expected games. Let* λ^2 *be the two-dimensional Lebesgue measure on* $\mathbb{R}^2$. *As the area of* W *is* $\frac{1}{2}$, *the density* f *of the uniform distribution* F *on* W *with respect to* λ^2 *is given by*

$$f(x,y) = \begin{cases} 2 & \text{if } (x,y) \in W \\ 0 & \text{otherwise.} \end{cases}$$

For symmetry reasons we have $E_v(\{1,3\}) = E_v(\{2,3\})$. *We obtain, using the density* f *and standard integration techniques,*

$$\begin{aligned} E_v(\{1,3\}) &= \int_W x \, dF(x,y) \\ &= \frac{1}{3}. \end{aligned}$$

For $E_{\hat{v}}$ *we have obviously* $E_{\hat{v}}(\{1,3\}) = E_{\hat{v}}(\{2,3\}) = \frac{1}{4}$. *On the set of known coalitions* $\mathcal{M}$ *the expected games equal the partially defined games. It is easy to see that the set of convex extensions of the sum of the two games* $v + \hat{v}$ *is the set* $W + \widehat{W}$. *Using that the area of* $W + \widehat{W}$ *is* $\frac{7}{4}$, *we obtain in the same way as for* E_v, *that the expected game* $E_{v+\hat{v}}$ *will satisfy*

$$\begin{aligned} E_{v+\hat{v}}(\{1,3\}) &= E_{v+\hat{v}}(\{2,3\}) \\ &= \frac{53}{84}. \end{aligned}$$

The Shapley values of the games $E_v, E_{\hat{v}}$ *and* $E_{v+\hat{v}}$ *are*

$$Sh(E_v) = \left(\frac{5}{18}, \frac{5}{18}, \frac{8}{18} \right)$$

$$Sh(E_{\hat{v}}) = \left(\frac{3}{8}, \frac{3}{8}, \frac{1}{4} \right) \text{ and}$$

$$Sh(E_{v+\hat{v}}) = \left(\frac{325}{504}, \frac{325}{504}, \frac{179}{252} \right).$$

So, $Sh(E_v) + Sh(E_{\hat{v}}) = \left(\frac{47}{72}, \frac{47}{72}, \frac{25}{36} \right) \neq Sh(E_{v+\hat{v}})$. $\Diamond$

We extend the model in such a way that players have a belief about the likelihoods of the underlying fully defined games. First of all we need to formalize the concept of a belief in this context. Let N be the set of players. The set G^N of fully defined games with player set N is isomorphic to $\mathbb{R}^{2^N \setminus \{\emptyset\}}$. Let $\mathcal{A}$ be the Lebesgue-Borel σ-algebra on $\mathbb{R}^{2^N \setminus \{\emptyset\}}$. This turns G^N into a measurable space $(G^N, \mathcal{A})$. The *belief* of the players can be expressed as a probability measure P over $(G^N, \mathcal{A})$. It is reasonable to restrict attention to those beliefs, of which the *support*

$$Supp(P) := \{ w \in G^N \mid P(A) > 0 \text{ for each open } A \ni w \}$$

contains only extensions of the game v. Furthermore, we assume that any coalition is expected to have a finite worth. For any $\mathcal{M}$-game v let Π^v be the set of all such probability measures P over $(G^N, \mathcal{A})$, i.e. that satisfy

$$Supp(P) \subseteq \{w \in G^N \mid w|_{\mathcal{M}} = v\}$$

and

$$-\infty < \int_{G^N} w(S) \, dP(w) < \infty \text{ for all } S \subseteq N.$$

We are now ready to define the set of all partially defined cooperative games *with beliefs* and player set N by

$$PGB^N := \{(v, P) \mid v \in PG^N \text{ and } P \in \Pi^v\}.$$

Let v be a fully defined game and for all $A \in \mathcal{A}$ let

$$\delta_v(A) = \begin{cases} 1 & \text{if } v \in A, \\ 0 & \text{otherwise.} \end{cases}$$

Then δ_v is the probability measure assigning probability 1 to the game v itself. Such measures are called *Dirac measures*. Note that $\Pi^v = \{\delta_v\}$ for any fully defined game. This embeds G^N into PGB^N by identifying w with (w, δ_w) for any fully defined cooperative game on N. Let

$$PGB := \bigcup\{PGB^N \mid \emptyset \neq N \subseteq \mathbb{N}, |N| < \infty\}$$

be the set of all partially defined cooperative games with beliefs. A *value solution* f for partially defined cooperative games with beliefs is a function

$$f : PGB^N \to \mathbb{R}^N.$$

4.3   Extending the Shapley value: The expected Shapley value

Next we present an extension of the Shapley value for fully defined games to partially defined games with beliefs. Young (1985) characterizes the Shapley value for fully defined games with efficiency, anonymity and strong monotonicity. We adapt these axioms to partially defined games with beliefs, and show that these adapted axioms give rise to a unique extension of the Shapley value to these games. As pointed out in Example 29, linearity, especially additivity, may be troublesome to use in a characterization of such an extension of the Shapley value. An adaptation of linearity is discussed later on in Section 4.4.

Let $f : PGB^N \to \mathbb{R}^N$ be a value solution. We call f *expectationally efficient* if it does not distribute more than what the grand coalition N will get in expectation, i.e.

$$\sum_{i \in N} f_i(v, P^v) = \int_{G^N} w(N) \, dP^v(w)$$

for any game (v, P^v) in PGB^N. If the worth of the grand coalition is not known, then *ex-post*, i.e. when a fully defined game $\hat{w}$ has been drawn from the carrier of P^v, ex-post efficiency

$$\sum_{i \in N} f_i(v, P^v) = \hat{w}(N)$$

can be violated. Players could insure against this with a risk-neutral insurer who also beliefs that the true game is drawn from P^v. It might also be that the game is played repeatedly and all realizations are independent . Then excesses and shortfalls of the worth of the grand coalition will cancel out in the long run. Then zero cost for insurance and equal interest rates to finance shortfalls and invest excesses have to be assumed.

If $N \in \mathcal{M}$ only extensions w where $w(N) = v(N)$ are in the carrier of the measure P^v, and the integral on the right hand side equals $v(N)$. So in that case expectational efficiency reduces to the equation

$$\sum_{i \in N} f_i(v, P^v) = v(N).$$

So we will have ex-post efficiency in that case.

It turns out that all we need for our characterization is *efficiency for fully defined games*, i.e.

$$\sum_{i \in N} f_i(v, P^v) = v(N),$$

for all $v \in PBG^N$ with coalition list $\mathcal{M} = 2^N$. Hence, accepting efficiency for fully defined games along with the other axioms used, implies acceptance of expectational efficiency.

Let $\tau : N \to N$ be a permutation of the player set N. For any fully defined game $w \in G^N$ the *permuted game of w*, w^τ is the fully defined game given by

$$w^\tau(S) := w(\tau^{-1}(S)) \text{ for all } S \subseteq N,$$

where

$$\tau^{-1}(S) := \{\tau^{-1}(i) \mid i \in S\}.$$

Let $A \in \mathcal{A}$, then the *permuted set A^τ* is given by

$$A^\tau := \{w^\tau \mid w \in A\}.$$

For the *permuted σ-algebra* $\mathcal{A}^\tau := \{A^\tau \mid A \in \mathcal{A}\}$ we obviously have $\mathcal{A}^\tau = \mathcal{A}$. So, if P is a probability measure on $(G^N, \mathcal{A})$ we can define the *permuted probability measure P^τ* on $(G^N, \mathcal{A})$ by

$$P^\tau(A) := P(A^{\tau^{-1}}) \text{ for all } A \in \mathcal{A}.$$

Furthermore, we can also permute partially defined cooperative games, but then the coali-
tion list has to be permuted also. So for any coalition list $\mathcal{M}$, let $\mathcal{M}^\tau$ be the *permuted
coalition list* given by

$$\mathcal{M}^\tau := \{\tau(S) \mid S \in \mathcal{M}\}.$$

If $\mathcal{M}$ is a symmetric coalition list, then $\mathcal{M}^\tau = \mathcal{M}$, and the set of known coalition sizes K
will not change. Let $v \in PG^N$ be a partially defined cooperative game with coalition list
$\mathcal{M}$, then the *permuted game* v^τ of v is the $\mathcal{M}^\tau$-game given by

$$v^\tau(S) := v(\tau^{-1}(S)) \text{ for all } S \in \mathcal{M}^\tau.$$

Finally, for any game $(v, P^v) \in PGB^N$, we call $(v^\tau, (P^v)^\tau)$ the permuted game of (v, P^v)
under τ. Obviously, we have $(v^\tau, (P^v)^\tau) \in PGB^N$.

   We call the value solution $f : PGB^N \to \mathbb{R}^N$ *anonymous* if

$$f((v^\tau, (P^v)^\tau)) = f^\tau(v, P^v) := (f_{\tau^{-1}(i)}(v, P^v))_{i \in N}$$

for any permutation τ of the player set and all $(v, P^v) \in PGB^N$. We call f *anonymous
for fully defined games* if we only demand this property to hold for fully defined games.
Note that $|\Pi^v| = |\{\delta_v\}| = 1$ for all such $v \in G^N$, so $f|_{G^N}$ does not depend on the belief.
Hence, anonymity for fully defined games is just the usual anonymity condition for a value
solution $g : G^N \to \mathbb{R}^N$, i.e. $g(w^\tau) = (g_{\tau^{-1}(i)}(w))_{i \in N}$ for all $w \in G^N$ and all permutations τ
of the player set N.

   In order to define strong monotonicity for $\mathcal{M}$-games with beliefs we use the *marginal
contributions operator* $\Delta_i(w) : 2^N \setminus \{\emptyset\} \to \mathbb{R}$ of a player $i \in N$ for fully defined games w
given by

$$\Delta_i(w)(S) := w(S \cup \{i\}) - w(S), \ S \in 2^N \setminus \{\emptyset\}.$$

For a game $(v, P^v) \in PGB^N$, and a player $i \in N$, let the *(expected) marginal contributions
operator* $\Delta_i(v, P^v) : 2^N \setminus \{\emptyset\} \to \mathbb{R}$ be given by

$$\Delta_i(v, P^v)(S) := \int_{G^N} \Delta_i(w)(S) dP^v(w).$$

The expected marginal contributions operator extends Δ_i to partially defined coopera-
tive games with beliefs. Using the (expected) marginal contributions operator for $\mathcal{M}$-
games with beliefs, we say that the value solution f is *strongly monotonic* if for all games
$(v, P^v), (\tilde{v}, P^{\tilde{v}}) \in PGB^N$ we have that if

$$\Delta_i(v, P^v) \geq \Delta_i(\tilde{v}, P^{\tilde{v}}),$$

for an $i \in N$, then

$$f_i(v, P^v) \geq f_i(\tilde{v}, P^{\tilde{v}}).$$

The *expected Shapley value* $Sh(v, P^v)$ of a partially defined game $(v, P^v) \in PGB^N$ is the
value solution given by

$$Sh(v, P^v) := \int_{G^N} Sh(w) \, dP^v(w).$$

Given any $\mathcal{M}$-game v with belief P^v we can define the expected game $\mathbf{E}_v$ of v under that belief by

$$\mathbf{E}_v(S) := \begin{cases} v(S) & \text{for all } S \in \mathcal{M}, \\ \int_{G^N} w(S)\, dP^v(w) & \text{otherwise.} \end{cases}$$

Then the expected Shapley value of (v, P^v) equals the Shapley value of the fully defined game $\mathbf{E}_v$. This follows by the linearity of the integral and the linearity of the Shapley value for fully defined games. Furthermore, linearity of the integral and of the marginal contributions operator for fully defined games imply that $\Delta_i(v, P^v)(S) = \Delta_i(\mathbf{E}_v)(S)$. We have $P^v = \delta_v$ for all $(v, P^v) \in G^N$ and so for fully defined games we obtain $v = \mathbf{E}_v$. Thus the expected Shapley value equals the Shapley value for fully defined games. Note that we write $Sh(v, P^v)$, $v \in PGB^N$ for the expected Shapley value, and $Sh(w)$, $w \in G^N$ for the Shapley value.

Theorem 30 *Let $f : PGB^N \to \mathbf{R}^N$ be a value solution that is efficient for fully defined games, anonymous for fully defined games, and strongly monotonic. Then f is the expected Shapley value.*

Proof. We use the characterization of Young (1985), who showed that any value solution for fully defined games satisfying efficiency, anonymity and strong monotonicity is the Shapley value on G^N. So

$$f|_{G^N} = Sh|_{G^N}. \tag{*}$$

Let $\mathbf{E}_v$ be the expected game of (v, P^v), $(v, P^v) \in PBG^N$. This defines a fully defined game. Let $i \in N$. As

$$\begin{aligned}
\Delta_i(v, P^v)(S) &= \int_{G^N} \Delta_i(w)(S)\, dP^v(w) \\
&= \int_{G^N} (w(S \cup \{i\}) - w(S))\, dP^v(w) \\
&= \int_{G^N} w(S \cup \{i\})\, dP^v(w) - \int_{G^N} w(S)\, dP^v(w) \\
&= \Delta_i(\mathbf{E}_v, \delta_{\mathbf{E}_v})(S),
\end{aligned}$$

we obtain by strong monotonicity and equation (*), that

$$\begin{aligned}
f_i(v, P^v) &= f_i(\mathbf{E}_v, \delta_{\mathbf{E}_v}) \\
&= Sh_i(\mathbf{E}_v).
\end{aligned}$$

As noted when defining the expected Shapley value we have by the linearity of the Shapley value and the linearity of the integral that

$$Sh_i(\mathbf{E}_v) = f_i(v, P^v).$$

∎

The expected Shapley value is not only efficient for fully defined games, but also expectationally efficient and anonymous for any partially defined cooperative game.

Lemma 31 *The expected Shapley value is expectationally efficient and anonymous.*

Proof. Let $(v, P^v) \in PGB^N$. We have $Sh_i(v, P^v) = Sh_i(\mathbb{E}_v)$ for all $i \in N$, and by expectational efficiency for fully defined games we obtain

$$\sum_{i \in N} Sh_i(v, P^v) = \sum_{i \in N} Sh_i(\mathbb{E}_v) = \mathbb{E}_v(N) = \int_{G^N} w(N) \, dP^v(w).$$

For anonymity, note that by anonymity for fully defined games, the substitution rule and $d(P^v)^\tau(w) = dP^v(w^{\tau^{-1}})$, we have that

$$
\begin{aligned}
Sh(v^\tau, (P^v)^\tau) &= \int_{G^N} Sh(w) d(P^v)^\tau(w) \\
&= \int_{G^N} Sh(w) dP^v(w^{\tau^{-1}}) \\
&= \int_{(G^N)^{\tau^{-1}}} Sh(w^\tau) dP^v(w) \\
&= \int_{G^N} Sh^\tau(w) dP^v(w) \\
&= Sh^\tau(v, P^v).
\end{aligned}
$$

■

4.4   Linearity of the expected Shapley value

In this section we shall define a weaker form of linearity, and give a characterization of the expected Shapley value with that weaker form of linearity, dummy player, efficiency and anonymity. First we note that partially defined games with the same coalition list can be added coalitionwise. The value solution $f : PGB^N \to \mathbb{R}^N$ is said to be *strongly linear* if for all $\alpha, \beta \in \mathbb{R}$ and $\mathcal{M}$-games $v, \hat{v}$ we have that

$$f(\alpha v + \beta \hat{v}) = \alpha f(v) + \beta f(\hat{v}).$$

Example 29 showed that the expected Shapley value does not satisfy strong linearity. However, taking beliefs into account may result in a linearity condition which is satisfied by the expected Shapley value. Moreover, we can define an addition operator on PGB^N also for games with different coalition lists.

For any $\alpha \in \mathbb{R}$ and any partially defined game (v, P^v) let $\overline{P}^{\alpha v}$ be the probability measure defined for $A \in \mathcal{A}$ by $\overline{P}^{\alpha v}(A) = P^v(\frac{1}{\alpha}A)$ if $\alpha \neq 0$, and $\overline{P}^{\alpha v} = \delta_{\underline{0}}$, if $\alpha = 0$, where $\underline{0}$ is the fully defined zero-game, in which all coalitions receive 0. For any two partially defined games $(v, P^v), (\hat{v}, P^{\hat{v}})$, let $\overline{P}^{v+\hat{v}}$ be the probability measure obtained by *convoluting* P^v and $P^{\hat{v}}$, i.e. for every $A \in \mathcal{A}$

$$\overline{P}^{v+\hat{v}}(A) = \int_{G^N} \int_{G^N} \mathbb{I}_A(x+y) dP^v(x) dP^{\hat{v}}(y),$$

where

$$I_A(x+y) = \begin{cases} 1 & \text{if } x+y \in A, \\ 0 & \text{otherwise.} \end{cases}$$

In words, $\overline{P}^{v+\widehat{v}}(A)$ is the probability that the sum of two extensions of the games (v, P^v) and $(\widehat{v}, P^{\widehat{v}})$ respectively is in A, given that the extensions are independently distributed according to P^v and $P^{\widehat{v}}$ respectively.

Let $(v, P^v), (\widehat{v}, P^{\widehat{v}}) \in PGB^N$ and $\alpha \in \mathbb{R}$. The games v and $\widehat{v}$ can be added coalitionwise where the worths of both games are known, i.e. on the coalition list $\mathcal{M} := dom(v) \cap dom(\widehat{v})$. We define an *addition* in PGB^N by

$$(v, P^v) \oplus (\widehat{v}, P^{\widehat{v}}) := \left((v|_{\mathcal{M}} + \widehat{v}|_{\mathcal{M}}), \overline{P}^{v+\widehat{v}} \right).$$

The *scalar multiplication* in PGB^N is defined by

$$\alpha \circledast (v, P^v) := (\alpha v, \overline{P}^{\alpha v}).$$

We call a value solution f *weakly linear* if

$$f(\alpha \circledast (v, P^v) \oplus \beta \circledast (\widehat{v}, P^{\widehat{v}})) = \alpha f(v, P^v) + \beta f(\widehat{v}, P^{\widehat{v}}),$$

for all $\alpha, \beta \in \mathbb{R}$ and $(v, P^v), (\widehat{v}, P^{\widehat{v}}) \in PGB^N$. So the additional constraint weakening linearity is on the belief for the game $\alpha v|_{\mathcal{M}} + \beta \widehat{v}|_{\mathcal{M}}$.

Next it is shown that the expected Shapley value satisfies weak linearity.

Theorem 32 *The expected Shapley value is weakly linear.*

Proof. Let $(v, P^v), (w, P^w) \in PGB^N$. The expected Shapley value satisfies

$$
\begin{aligned}
Sh((v, P^v) \oplus (w, P^w)) &= \int_{G^N} Sh(x) \, d\overline{P}^{v+w}(x) \\
&= \int_{G^N} \int_{G^N} Sh(x - y + y) \, dP^v(x-y) \, dP^w(y) \\
&= \int_{G^N} \int_{G^N} (Sh(x-y) + Sh(y)) \, dP^v(x-y) \, dP^w(y) \\
&= \int_{G^N} \left[\left(\int_{G^N - t} Sh(\overline{x}) \, dP^v(\overline{x}) \right) + Sh(y) \right] dP^w(y) \\
&= \int_{G^N} \left[\left(\int_{G^N} Sh(\overline{x}) \, dP^v(\overline{x}) \right) + Sh(y) \right] dP^w(y) \\
&= \int_{G^N} [Sh(v, P^v) + Sh(y)] \, dP^w(y) \\
&= Sh(v, P^v) \, P^w(G^N) + \int_{G^N} Sh(y) \, dP^w(y) \\
&= Sh(v, P^v) + Sh(w, P^w).
\end{aligned}
$$

Let $\alpha \in \mathbb{R}$. We are done if we can show that the expected Shapley value also satisfies $Sh(\alpha \circledast (v, P^v)) = \alpha Sh(v, P^v)$. This follows as for $\alpha \neq 0$ we have

$$
\begin{aligned}
Sh(\alpha \circledast (v, P^v)) &= \int_{G^N} x \, d\overline{P}^{\alpha v}(x) \\
&= \int_{G^N} x \, dP^v(\frac{1}{\alpha}x) \\
&= \int_{\frac{1}{\alpha}G^N} \alpha x \, dP^v(x) \\
&= \alpha \int_{G^N} x \, dP^v(x) \\
&= \alpha Sh(v, P^v),
\end{aligned}
$$

and for $\alpha = 0$ we have

$$
Sh(\alpha \circledast (v, P^v)) = Sh(\underline{0}, \delta_{\underline{0}}) = (0, \ldots, 0) = 0 Sh(v, P^v).
$$

∎

The expected Shapley value can also be characterized with weak linearity. One of the standard axioms that we will need is the null player axiom. Let $(v, P^v) \in PGB^N$. We call a player $i \in N$ a *null player* in (v, P^v) if the (expected) marginal contributions of that player are 0, i.e. if

$$
\Delta_i(v, P^v)(S) = 0 \text{ for all } S \subseteq N.
$$

We say that the value solution $f : PGB^N \rightarrow \mathbb{R}^N$ satisfies the *null player property*, if for all $(v, P^v) \in PGB^N$, and all null players i in the game (v, P^v), we have

$$
f_i(v, P^v) = 0.
$$

Theorem 33 *Let $f : PGB^N \rightarrow \mathbb{R}^N$ be a value solution satisfying efficiency, anonymity, null player and weak linearity. Then f is the expected Shapley value.*

Proof. For fully defined games weak linearity and strong linearity are the same. Shapley (1953) has characterized the Shapley value for fully defined games with efficiency, anonymity, null player and strong linearity. So

$$
f|_{G^N} = Sh|_{G^N}. \tag{*}
$$

Let (v, P^v) be a partially defined cooperative game with coalition list $\mathcal{M}$. Let δ_{E_v} be the Dirac measure on the expected game E_v of (v, P^v). The expected game E_z of the game

$$
(z, P^z) := (v, P^v) \oplus ((-1) \otimes (E_v, \delta_{E_v}))
$$

satisfies

$$
\begin{aligned}
\mathbf{E}_z(S) &= \int_{G^N} \int_{G^N} (x+y)\, dP^v(x)\, d(\delta_{\mathbf{E}_v})(-y) \\
&= \int_{G^N} \int_{G^N} (x-\tilde{y})\, dP^v(x)\, d(\delta_{\mathbf{E}_v})(\tilde{y}) \\
&= \int_{G^N} (\mathbf{E}_v - y)\, d(\delta_{\mathbf{E}_v})(y) \\
&= 0
\end{aligned}
$$

for all $S \in 2^N$. As $\Delta_i(v, P^v) = \Delta_i(\mathbf{E}_z)$ this shows that all players in (z, P^z) are null players. Hence,

$$
f(z, P^z) = (0, \ldots, 0).
$$

As (z, P^z) was defined to be

$$
((v, P^v) \oplus (-1) \circledast (\mathbf{E}_v, \delta_{\mathbf{E}_v}))
$$

this implies by weak linearity that

$$
f(v, P^v) - f(\mathbf{E}_v, \delta_{\mathbf{E}_v}) = (0, \ldots, 0).
$$

As $(\mathbf{E}_v, \delta_{\mathbf{E}_v})$ can be identified with the fully defined game $\mathbf{E}_v \in G^N$ we have proved using (∗) that

$$
f(v, P^v) = Sh(\mathbf{E}_v, \delta_{\mathbf{E}_v}).
$$

Again, $Sh(\mathbf{E}_v, \delta_{\mathbf{E}_v})$ equals the expected Shapley value $Sh(v, P^v)$. ∎

The proof uses efficiency and anonymity only on fully defined games. Weak linearity can be weakened to weak additivity, just as in the corresponding theorem of Shapley (1953).

4.5 Representation, Dividends and Potential

Some important results for fully defined games concern representation by unanimity games, dividend and potential. This section will show how to extend these results to partially defined cooperative games with beliefs.

First we look at the representation by unanimity games. For nonempty subsets T of N the *T-unanimity game* is a fully defined game, say u_T, defined by

$$
u_T(S) = \begin{cases} 1 & \text{if } T \subseteq S, \\ 0 & \text{otherwise.} \end{cases}
$$

For any fully defined game w we have the representation

$$
w = \sum_{T \in 2^N \setminus \{\emptyset\}} c_T u_T,
$$

where

$$c_T = \sum_{S:S \subseteq T} (-1)^{|T|-|S|} w(S).$$

The set $\{u_T \mid T \in 2^N \setminus \{\emptyset\}\}$ is linearly independent, so it forms a basis for the set of fully defined games. The question arises whether a representation with unanimity games is also possible for partially defined games with beliefs.

Let $C : G^N \to \mathbb{R}^{2^N \setminus \{\emptyset\}}$ be the mapping given by $C(w) = (c_T)_{T \in 2^N \setminus \{\emptyset\}}$, whenever $w = \sum_{T \in 2^N \setminus \{\emptyset\}} c_T u_T$. Then C is bijective and measurable as a mapping from $(G^N, \mathcal{A})$ to $(\mathbb{R}^{2^N \setminus \{\emptyset\}}, \mathcal{A})$. Let (v, P^v) be a partially defined cooperative game with beliefs. The probability measure P^C on $(\mathbb{R}^{2^N \setminus \{\emptyset\}}, \mathcal{A})$ given by $P^C := P^v \circ C^{-1}$, represents P^v in the following sense:

$$
\begin{aligned}
P^v(A) &= (P^v \circ C^{-1} \circ C)(A) \\
&= P^C(C(A)) \text{ for all } A \in \mathcal{A}.
\end{aligned}
$$

So, beliefs about the underlying fully defined game can also be seen as beliefs about possible representations of these underlying games by unanimity games.

Next we want to extend the concept of a dividend to partially defined cooperative games with beliefs. Let w be a fully defined game. For each coalition T, Harsanyi (1959) defined a real number d_T, the *dividend*, in a recursive manner as follows.

$$
d_T(w) := \begin{cases} 0 & \text{if } T = \emptyset, \\ |T|^{-1} \left(w(T) - \sum_{S \subsetneq T} |S| d_S(w) \right) & \text{if } |T| \geq 1. \end{cases}
$$

If $w = \sum c_T u_T$ then $|T| d_T(w) = c_T$ for all $T \neq \emptyset$ and $Sh_i(w) = \sum_{T:i \in T} d_T(w)$, i.e. the Shapley value of a player is the sum of all dividends of coalitions to which the player belongs.

Taking into account the beliefs of the players about the underlying games, will turn dividends into random variables, with a probability measure depending on the beliefs of the players. Let (v, P^v) be a partially defined cooperative game. Let $C : G^N \to \mathbb{R}^{2^N \setminus \{\emptyset\}}$ and P^C be as defined above. Let $D : G^N \to \mathbb{R}^{2^N \setminus \{\emptyset\}}$ be the mapping given by $D(w) = (d_T(w))_T$. Then D is bijective and measurable as a mapping from $(G^N, \mathcal{A})$ to $(\mathbb{R}^{2^N \setminus \{\emptyset\}}, \mathcal{A})$. Defining P^D as the probability measure on $(\mathbb{R}^{2^N \setminus \{\emptyset\}}, \mathcal{A})$ such that $P^D := P^v \circ D^{-1}$ gives the desired analogy for the *random dividend* $D : (G^N, \mathcal{A}) \to (\mathbb{R}^{2^N \setminus \{\emptyset\}}, \mathcal{A})$ distributed according to P^D. As the reader can verify using the corresponding facts for fully defined games we have

1. the mapping $F : \mathbb{R}^{2^N \setminus \{\emptyset\}} \to \mathbb{R}^{2^N \setminus \{\emptyset\}}$ given by $F(d) := (|T| d_T)_{T \in \mathbb{R}^{2^N \setminus \{\emptyset\}}}$ is bijective and measurable as a mapping from $(\mathbb{R}^{2^N \setminus \{\emptyset\}}, \mathcal{A})$ to $(\mathbb{R}^{2^N \setminus \{\emptyset\}}, \mathcal{A})$.

2. $F(D) = C$, i.e. $P^D(A) = P^C(F(A))$ for any $A \in \mathcal{A}$.

3.

$$Sh_i(v, P^v) = \sum_{T:i\in T} \int_{G^N} (D(w))_T \, dP^v(w)$$

$$= \sum_{T:i\in T} \int_R d_T \, dP^{D_T}(d_T)$$

where the measure P^{D_T} on R is the image of the measure P^D under the projection on the coordinate (coalition) T, i.e.

$$P^{D_T}(B) = P^D(\{x \in \mathbb{R}^{2^N} \mid x_T \in B\})$$

for all Lebesgue-Borel measurable sets $B \subseteq R$.

The last extension concerns the potential of a game. A *potential* (Hart and Mas-Colell, 1989) is a real valued function $Q : G \to \mathbb{R}$ such that

$$Q(\emptyset, w) = 0$$
$$\sum_{i\in N} D_i Q(N, w) = w(N) \text{ for all } w \in G^N. \quad (\text{Q-efficiency})$$

Here $D_i Q(N, w) = Q(N, w) - Q(N \setminus \{i\}, w|_{N\setminus\{i\}})$. If Q is a potential then the *gradient* $gradQ(N, w) := (D_i Q(N, w))_{i\in N}$ is an efficient payoff vector for the game. By Q-efficiency

$$Q(N, w) = \frac{1}{|N|} \left(w(N) + \sum_{i\in N} Q(N \setminus \{i\}, w|_{N\setminus\{i\}}) \right).$$

So the potential of $w \in G^N$ can be determined by the potentials of the subgames. By $Q(\emptyset, w) = 0$ it is therefore uniquely defined. The potential is connected to the dividends and the Shapley value of fully defined games in the following way (Hart and Mas-Colell, 1989):

$$Q(N, w) = \sum_{\emptyset \neq T \subseteq N} d_T$$
$$gradQ(N, w) = Sh(w).$$

Let (v, P^v) be a partially defined cooperative game with player set N and coalition list $\mathcal{M}$. The *restriction* of w to the player set $\tilde{N} \subseteq N$ is the partially defined cooperative game $(v|_{\tilde{N}}, P^{v|\tilde{N}})$ with player set $\tilde{N}$, coalition list $\tilde{\mathcal{M}} = \mathcal{M} \cap 2^{\tilde{N}}$ and belief given by

$$P^{v|\tilde{N}}(A) = P^v(\{x \in \mathbb{R}^{2^N\setminus\{\emptyset\}} \mid x|_{\tilde{\mathcal{M}}} \in A\})$$

for all A in the Lebesgue-Borel σ-algebra on $\mathbb{R}^{2^{\tilde{N}}\setminus\{\emptyset\}}$. A *potential* on the set of partially defined cooperative games with beliefs is a real valued function $Q : PGB \to \mathbb{R}$. The result for fully defined games can be extended to hold also for partially defined games when requiring instead of Q-efficiency

$$\sum_{i\in N} D_i Q(N, v, P^v) = \int_{G^N} w(N) dP^v(w) \quad (\text{expected Q-efficiency})$$

for all partially defined games (v, P^v) with player set N. Define $D_i Q(N, v, P^v)$, $gradQ(N, v, P^v)$ just as for fully defined games. The gradient gives us by expectational Q-efficiency an expectationally efficient payoff vector. Furthermore, the potential is uniquely defined, extends the potential for fully defined games and can be connected to the expected Shapley value and expectations of the random dividends for partially defined games in a similar manner as for fully defined games.

4.6   Reduced and normalized Shapley Value

This section shows how to embed the reduced Shapley value (Willson, 1993) and the normalized Shapley value (Housman, 2001) into our model. Both these values are only defined for symmetric coalition lists containing the grand coalition, and do not depend on the belief. So we can restrict attention to

$$\widehat{PG}^N := \{v \in PG^N \mid dom(v) \text{ is a symmetric coalition list, } N \in dom(v)\}.$$

Let $n := |N|$ be the cardinality of the player set N. The reduced Shapley value of a game v in $\widehat{PG}^N$ is given by

$$RSh_i(v) := \frac{1}{n} \sum_{m \in K} \left(\binom{n-1}{m-1}^{-1} \sum_{\substack{|S|=m \\ i \in S}} v(S) - \binom{n-1}{m}^{-1} \sum_{\substack{|S|=m \\ i \notin S}} v(S) \right)$$

for any player $i \in N$, where K are the known coalition sizes. This can also be written in terms of marginal (known) contributions. Let

$$\rho(m) := max\{k \in K \mid k < m\}$$

be the largest known coalition size smaller than $m \in K$. The marginal known contributions of a player $i \in N$ to a coalition $S \ni i$, $|S| \in K$, are then given by $w(S) - w(R)$ for any $R \subseteq S - \{i\}$, $|R| = \rho(|S|)$. Willson (1993) characterizes the reduced Shapley value by a strong monotonicity condition based on these marginal contributions, together with efficiency, strong anonymity and strong linearity. Housman (2001) notes that it can also be expressed using these marginal contributions as

$$RSh_i(v) = \frac{1}{n} \sum_{m \in K} \binom{n-1}{m-1}^{-1} \sum_{\substack{|S|=m \\ i \in S}} \binom{m-1}{\rho(m)} \sum_{\substack{|R|=\rho(m) \\ R \subseteq S - \{i\}}} (w(S) - w(R))$$

for any player $i \in N$. Willson (1993) and Housman (2001) also note that the reduced Shapley value of $v \in \widehat{PG}^N$ equals the Shapley value of the extension w, where the unknown coalition worths are set equal to a constant which is allowed to depend on the size, i.e. $w|_{\mathcal{M}} = v$ and $w(S) := c_{|S|}$ for all $S \notin \mathcal{M}$, where $c_k \in \mathbb{R}$ for all $k \in \{0, \ldots, n\} \setminus K$. The

Shapley value will then be independent of the chosen constants, the corresponding terms cancel out and only the terms forming the reduced Shapley value influence the value. Using that, we can express the reduced Shapley value as an expected Shapley value.

Lemma 34 *Let* $v \in \widehat{PG}^N$. *Let* $P^v \in \Pi^v$ *be a belief leading to an expected game* $\mathbf{E}_v$ *satisfying*

$$\mathbf{E}_v(S) = \mathbf{E}_v(T)$$

for all $S, T \subseteq N$ *such that* $|S| = |T| \in \{0, \dots, n\} \setminus K$. *Then* $Sh(v, P^v) = RSh(v)$.

Proof. This is just another way of saying that the unknown coalition worths are equal to a size-dependent constant. ∎

It is not hard to find beliefs satisfying this lemma. Just take the Dirac measure on an extension where coalitions of the same size have the same value if their worths are unknown. The lemma says that the reduced Shapley value amounts to assuming implicitly that players believe that extensions, which are constant across unknown worths of coalitions of the same size, give good estimates of the missing values. That this can be problematic was already known to Willson (1993), who notes that usually such extensions will not satisfy properties such as monotonicity or superadditivity. Housman (2001) gives an example where the reduced Shapley value of a game is not among the Shapley values of any of its superadditive extensions. So, if players have such beliefs about the underlying game it might not be reasonable to use the reduced Shapley value.

Housman (2001) defined the normalized Shapley value for player $i \in N$ by

$$NSh_i(v) = \frac{1}{n} \sum_{m \in K} \left(\binom{n-1}{m-1}^{-1} \sum_{\substack{|S|=m \\ i \in S}} v(S) - \binom{n-1}{m}^{-1} \sum_{\substack{|S|=m \\ i \notin S}} v(S) \right)$$
$$+ \left(1 - \frac{|K|}{n} \right) \left(v(i) - \frac{1}{n-1} \sum_{j \in N \setminus \{i\}} v(j) \right),$$

as an alternative if $1 \in K$. If $1 \notin K$ then the normalized Shapley value is defined to equal the reduced Shapley value, i.e. the last line of the sum disappears. The following lemma shows that the normalized Shapley value can also be obtained as an expected Shapley value, and to which estimates of the missing worths this amounts. It is stated only for symmetric coalition lists where $1 \in K$ (individual worths are known), as the normalized Shapley value equals the reduced Shapley value for games where $1 \notin K$.

Lemma 35 *Let* $v \in \widehat{PG}^N$ *and suppose that* $1 \in K$. *Let* $P^v \in \Pi^v$ *be a belief leading to an expected game* $\mathbf{E}_v$ *satisfying*

$$\mathbf{E}_v(S) = \sum_{i \in S} v(\{i\})$$

for all S *such that* $S \notin \mathcal{M}$. *Then* $Sh(v, P^v) = NSh(v)$.

Proof. Straightforward but tedious. ∎

Using Dirac measures as beliefs it is not hard to specify beliefs satisfying the lemma. But again the expected games will usually violate common assumptions players might have. However, in a sense the lemma reveals that these estimates are a bit more sophisticated. They take into account at least the individual worths of the players, will in general not be constant across coalitions of the same size, and grow with coalition size if individual worths are positive. This method is appropriate for example if players believe they are playing additive games, and could probably still be used for beliefs, which are 'not too far away' from the belief to play an additive game. However, there is no intuitive reason why these worths should be used in every case. One could imagine many other, even more sophisticated estimates, e.g.

$$w(S) = max\{\sum_{T \in \mathcal{P}} v(T) \mid \mathcal{P} \text{ any partition of } S \text{ into sets } T, T \in \mathcal{M}\},$$

$$w(S) = min\{\sum_{T \in \mathcal{P}} v(T) \mid \mathcal{P} \text{ any partition of } S \text{ into sets } T, T \in \mathcal{M}\},$$

or $w(S)$ being any weighted average of the sum of coalition worths in such partitions. All of these can be accommodated as expected Shapley values with respect to specific beliefs. Whether there are axioms characterizing the corresponding Shapley values on $\widehat{PG}^N$ in the spirit of Willson (1993) and Housman (2001), i.e. without beliefs, is an open problem.

As mentioned already, Housman (2001) gave an example where the reduced Shapley value does not equal the Shapley value of any of the superadditive extensions of the partially defined cooperative game under consideration. It is not hard to imagine that the same problem arises with the normalized Shapley value. There the extension which is implicitly assumed to be the expected game might not be superadditive. The expected Shapley value behaves better in that sense, i.e. it picks a value from the set of values of the extensions, if one imposes a restriction on the support of the beliefs.

Lemma 36 *Let (v, P^v) be a partially defined cooperative game. Suppose that the support of P^v is convex. Then the expected Shapley value of (v, P^v) is equal to the Shapley value of an extension $w \in G^N$ of v, considered possible by the players, i.e.*

$$Sh(v, P^v) \in \{Sh(w) \mid w \in Supp(P^v)\}.$$

Proof. As the support $Supp(P^v)$ of P^v is convex we have $\mathbf{E}_v \in Supp(P^v)$. Hence, $Sh(v, P^v) = Sh(\mathbf{E}_v) \in \{Sh(w) \mid w \in Supp(P^v)\}$. ∎

This restriction is not unnatural and satisfied in important situations, such as monotonic, superadditive and convex extensions, or any other set of extensions arising from l homogeneous linear inequalities,

$$Hw \geq 0, \ H \in \mathbb{R}^{l \times 2^N \setminus \{\emptyset\}}, l \text{ any nonnegative integer,}$$

on the vector $(w(S))_{S \in 2^N \setminus \{\emptyset\}}$ of coalition worths.

4.7   Conclusion

Adding belief to Willson's (1993) and Housman's (2001) definitions on partially defined cooperative games was shown to evolve into the existence of a natural and unique extension of the Shapley value to such games. It is defined for any coalition list and any belief having finite expectations of coalitional worths. It can be characterized by properties analogous to these of Young (1985) or Shapley (1953). Representation with unanimity games, dividend and potential can be extended to partially defined cooperative games with beliefs. The reduced Shapley value and the normalized Shapley can be incorporated as special cases. Housman (2001) already concluded, future research might consider situations where players have the option to choose among several coalition lists. This might be the case in practise, if players can, due to cost and complexity, only determine a limited number of worths. In that case it is of interest to investigate which worths should be determined. Another topic could be to extend the core to partially defined cooperative games.

Minimal manipulability of social

choice functions

4.7 Conclusion

Adding belief to Wilkes's (1990) and Housman's (2001) definitions on partially defined cooperative games was shown to evolve into the existence of a natural and unique extension of the Shapley value to such games. It is defined for any coalition list and any belief has on finite coalitions of coalitional worths. It can be characterized by properties analogous to those of Young (1985) or Shapley (1953). Representation with quasiminity games, dividend and potential can be extended to partially defined cooperative games with beliefs. The reduced Shapley value and the normalized Shapley can be incorporated as special cases. Housman (2001) already concluded, future research might consider situations where players have the option to choose among several coalition lists. This might be the case in practice if players can, due to cost and complexity, only determine a limited number of worths. In that case it is of interest to investigate which worth should be determined. Another topic could be to extend the one to partially defined cooperative games

Proof. Straightforward but tedious.

Part III

Minimal manipulability of social choice functions

Part III

Minimal manipulability of social choice functions

Chapter 5

Unanimity and Nondictatorship

This chapter is concerned with the number of profiles at which a nondictatorial social choice function is manipulable. For three or more alternatives the lower bound is derived when the social choice function is nondictatorial and unanimous. In the case of three alternatives the lower bound is also derived when the social choice function is nondictatorial and surjective. In both cases all social choice functions reaching that lower bound are characterized when there are at least three agents. In the case of two agents the characterized social choice functions are only a subset of the set of all social choice functions reaching the minimum.[1]

5.1 Introduction

A well-known result of Gibbard (1973) and Satterthwaite (1975) shows that any surjective nondictatorial social choice function with more than two alternatives must be manipulable. However, little is known about the degree of manipulability of nondictatorial social choice functions. An investigation was pioneered by Kelly (1988), who gave the minimal number of manipulable profiles for social choice functions with three alternatives and two agents, and formulated several conjectures about the general case. This line of research was continued by Fristrup and Keiding (1998), who gave the minimal number of manipulable profiles for two agents and any number of alternatives. It was also conjectured in their paper, that there is hope that the social choice functions they use also give the minimum in the general case.

We prove their conjecture if surjectivity is replaced by unanimity, but show that it is not true with only surjectivity in the case of three alternatives. The minimally manipulable social choice functions given as examples in Fristrup and Keiding (1998) can be described as almost dictatorial social choice functions. We show that these form the set of all minimally manipulable nondictatorial and unanimous social choice functions in the case of three agents or more, and a subset of this set in the case of two agents. Example 4 in Kelly (1989) shows a unanimous two person social choice function hitting the minimum, but

---

[1]The results in this chapter were first formulated in Maus et al. (2004b).

which is not almost dictatorial. In the case of three alternatives almost dictatorial social choice functions lead to $2n - 1$ manipulable profiles where n is the number of agents and there are at least three agents. We present six social choice functions that are manipulable at n profiles. We characterize these social choice functions as all minimally manipulable social choice functions for three alternatives and more than two agents. For $n = 2$ these social choice functions reach the minimum of 2 manipulable profiles, but there are many other such social choice functions that do this, again see e.g. Example 4 in Kelly (1989). We mention that an optimization program we wrote shows that there are 135 social choice functions reaching the minimum of two manipulable profiles in the case of two agents and three alternatives.

Another surprising feature of the minimally manipulable social choice functions characterized in the three alternative surjectivity case is that they are anonymous. Considering the distribution of power among agents, anonymous social choice functions are rather far away from dictatorial or almost dictatorial social choice functions. However, we did not find a generalization of these social choice functions to more than three alternatives that beats almost dictatorial social choice functions in that case. In fact, in Maus et al. (2004b, see Chapter 6) we show that any surjective and anonymous social choice function has more manipulable profiles than an almost dictatorial one if $m \geq 4$ and $n \geq m + 2$, where m is the number of alternatives. So the conjecture of Fristrup and Keiding (1998) can still be true for more than three alternatives, and as mentioned above we show that it is true if one replaces surjectivity by unanimity.

First, Section 5.2 is on notation and model description. Section 5.3 introduces the concept of manipulation and some basics about it. After that Section 5.4 concerns wellknown results on minimal manipulability. Section 5.5 contains the result for unanimous and nondictatorial social choice functions, and Section 5.6 contains the result for surjective and nondictatorial social choice functions with three alternatives. Finally a conclusion is given in Section 5.7. The appendix is on unanimous and nondictatorial social choice functions for the special three agents case.

5.2   Preliminaries

We denote the cardinality of a set S by $|S|$ and its powerset by 2^N.

Let A be a finite set of alternatives, $m := |A| \geq 3$, and N a finite set of agents, $n := |N| \geq 2$.

Let $t \subset A \times A$. We call t *complete* if for all $x, y \in A$ $(x, y) \in t$ or $(y, x) \in t$. Note that completeness of t implies $(x, x) \in t$ for all $x \in A$. We call t *transitive* if for all $x, y, z \in A$ $(x, y) \in t$ and $(y, z) \in t$ implies $(x, z) \in A$. We call t *antisymmetric* if for all $x, y \in A$ $(x, y) \in t$ and $(y, x) \in t$ implies that $x = y$.

A *preference* $t \subset A \times A$ is a *linear ordering* (complete, transitive, antisymmetric) on A. Let P denote the set of all preferences. Suppose that $A = \{x_1, x_2, \ldots x_m\}$. By completeness,

transitivity and antisymmetry we can write conveniently

$$t = x_1 x_2 \dots x_m$$

for the preference t such that $(x_i, x_j) \in t$ if and only if $i \geq j$, $i, j \in \{1, 2, \dots, m\}$,

$$t = \dots x \dots y \dots$$

if we want to express only that x is strictly preferred to y, and

$$t = x \dots$$

if we want to express only that x is preferred to all other alternatives. Furthermore we use

$$\dots x \dots y \dots z \dots, \quad xy \dots \text{ and } xyz.$$

whose meanings can be easily deduced.

A *profile* p is a map from N to P. Let P^N denote the set of all these maps. Thus, a profile assigns to every agent i a preference $p(i)$ over the alternatives. For a nonempty subset S of N we denote by $p|_S$ the restriction of the map p to the domain S. By $(p|_{N-S}, t^S)$ we denote the profile q such that

$$q(i) := \begin{cases} p(i) & \text{for all } i \in N - S, \\ t & \text{for all } i \in S. \end{cases}$$

In the particular case $S = \{i\}$ we write (p_{-i}, t) instead of $(p_{N-\{i\}}, t^{\{i\}})$, and if $p(N - \{i\}) = \{\tilde{t}\}$ we write $(\tilde{t}^{N-\{i\}}, t)$. For a profile $p \in P^N$ and alternatives $x, y \in A$ let $S^{(x,y)}(p) := \{i \in N \mid (x, y) \in p(i)\}$ be the set of all agents that prefer x to y.

For two profiles $p, q \in P^N$ we define the *distance* between p and q by $dist(p, q) := |\{i \in N \mid p(i) \neq q(i)\}|$. A finite sequence $r^0, \dots, r^l \in P$ such that $f(r^0) = p$, $f(r^l) = q$, and $dist(r^i, r^{i+1}) = 1$ for all $i = 0, \dots, l - 1$ is called a *path* from p to q. A path $r^0, \dots, r^l$ such that $l = dist(p, q)$ is called a *shortest* path from p to q.

For a profile p and a set of profiles Q we define the *distance* between p and Q by $dist(p, Q) := min_{q \in Q} dist(p, q)$.

A *social choice function* is a function $f : P^N \rightarrow A$. Hence, a social choice function selects a unique alternative $f(p)$ at every profile p.

A social choice function is called *surjective* if any alternative in A is chosen at least once, i.e. if $f(P^N) = A$. In literature this is also known as citizen-sovereignty. A social choice function is called *unanimous* if $f(p) = x$ for all profiles $p \in P^N$ such that $p(i) = x \dots$ for all $i \in N$, i.e. for all profiles in $top(x) := \{p \in P^N \mid p(i) = x \dots \text{ for all } i \in N\}$. Clearly, unanimity is stronger than surjectivity. Throughout this chapter we assume that any social choice function is at least surjective.

For a permutation σ of N and a profile $p \in P^N$ let $p \circ \sigma$ be the profile given by $(p \circ \sigma)(i) := p(\sigma(i))$ for all $i \in N$. A social choice function is called *anonymous* if $f(p) = f(p \circ \sigma)$ for

all permutations σ of N. Thus, anonymous social choice functions are symmetric in the arguments. In a sense they treat agents equally.

In contrast to anonymity, the following dictatorial social choice functions $dict_d$ respect only the preference of one single agent $d \in N$, the *dictator*. For any profile p $dict_d$ is defined by

$$dict_d(p) := x$$

where x is such that $p(d) = x \ldots$ So, $dict_d(p)$ is the most preferred alternative of agent d in $p(d)$. A social choice function f is called *nondictatorial* if there is no agent d such that $f = dict_d$.

5.3   Manipulation of social choice functions

We are interested in strategic behaviour of individuals when facing cooperative decision-making as captured by social choice functions. This is formalised by the following definitions. Let f be a social choice function.

A social choice function is said to be *intermediate manipulable* at a profile p (by coalition $S \subseteq N$, $S \neq \emptyset$), if $p(S) = \{t\}$ for some $t \in P$, and there is an *S-deviation* q, i.e. $q|_{N-S} = p|_{N-S}$ and $q(S) = \{\widehat{t}\}$, such that $(f(p), f(q)) \notin t$. We call p *intermediate manipulable towards q*. Let

$$IM_f := \{p \in P^N \mid f \text{ is intermediate manipulable at profile } p\}.$$

An S-deviation where $S = \{i\}$, $i \in N$, is called an *i-deviation*. If it is not important which set of agent deviates from p to q we call q a *deviation* from p. A profile p is called *(individually) manipulable* (under f) if there is an agent that is better off by being dishonest about his preference, i.e. if there is an $i \in N$, and an i-deviation q such that

$$(f(p), f(q)) \notin p(i).$$

We call p *manipulable towards q*. Let

$$M_f := \{p \in P^N \mid p \text{ is manipulable under } f\}.$$

A social choice function is called *strategy-proof* if $M_f = \emptyset$, otherwise it is said to be *(individually) manipulable*. Note that $M_f = \emptyset$ for all dictatorial rules f. The prominence of the dictatorial rules arises from the following impossibility result due to Gibbard (1973) and Satterthwaite (1975).

Theorem 37 *Let A be a finite set of alternatives, $|A| \geq 3$. Let $f : P^N \to A$ be a nondictatorial surjective social choice function. Then*

$$|M_f| \geq 1.$$

We show some useful connections between intermediate manipulability and manipulability. Clearly, $M_f \subseteq IM_f$. If $p \in IM_f - M_f$ is intermediate manipulable towards q, the following (standard) lemma holds.

Lemma 38 *Let* $f : P^N \to A$. *Let* p *be intermediate manipulable towards* q, *but not manipulable. Let* $r^0, \ldots, r^l$ *be a shortest path from* p *to* q. *Then there is a* $k \in \{1, \ldots, l-1\}$ *such that* r^k *is manipulable towards* r^{k+1}.

Proof. Let $D = \{i \in N \mid p(i) \neq q(i)\}$. Since f is intermediate manipulable at p by D towards q, we have $p(D) = \{t\}$ for some $t \in P$, so $p = (t^D, p|_{N-D})$ and $(f(p), f(q)) \notin t$. As $r^0, \ldots, r^l$ is a shortest path from p to q, r^{k+1} is a j-deviation from r^k for every $k \in \{0, \ldots, l\}$ for some agent $j \in D$, and $r^k(j) = t$. By the transitivity of t and $(f(r^0), f(r^l)) \notin t = r^k(j)$, there must be at least one $k \in \{0, \ldots, l-1\}$ such that $(f(r^k), f(r^{k+1})) \notin t$. Then r^k is manipulable towards r^{k+1}. As $r^0 \notin M_f$, $k \geq 1$. ∎

Each set of disjoint shortest paths from p to q contains at most $dist(p,q)$ elements. So we can state the following corollary to Lemma 38.

Corollary 39 *Let* $f : P^N \to A$. *Let* p *be intermediate manipulable towards* q, *but not manipulable. Then there are at least* $dist(p,q)$ *profiles in* M_f, *each of them on a shortest path from* p *to* q.

Let $p, q \in IM_f - M_f$ be such that p is intermediate manipulable towards q and vice versa. Suppose that there is a shortest path from p to q that contains only one manipulable profile. Then we can state the following about the preferences of the manipulating coalition S in p and q.

Lemma 40 *Let* $p, q \in IM_f - M_f$ *be such that* p *is intermediate manipulable by* S *towards* q *and* q *is intermediate manipulable by* S *towards* p. *Suppose that there is a shortest path* $r^0, \ldots, r^l$ *from* p *to* q *such that* $\{r^0, \ldots, r^l\} \cap M_f = \{r^k\}$. *Let* $x := f(p)$, $y := f(q)$, $z := f(r^k)$. *Then* $z \in A - \{x, y\}$ *and* $p(S) = \{t\}, q(S) = \{\tilde{t}\}$, *are such that* $t = \ldots y \ldots x \ldots z \ldots$ *and* $\tilde{t} = \ldots x \ldots y \ldots z \ldots$. *In particular, if* $m = 3$ *then* $t = yxz, \tilde{t} = xyz$, *and* $f(r^i) = x$ *for all* $i \in \{0, \ldots, k-1\}$, $f(r^i) = y$ *for all* $i \in \{k+1, \ldots, l\}$.

Proof. Note that $r^0, \ldots, r^l$ is a shortest path from p to q and $r^l, \ldots, r^0$ is a shortest path from q to p. By the intermediate manipulability of p towards q and q towards p we have $t = \ldots y \ldots x \ldots$ and $\tilde{t} = \ldots x \ldots y \ldots$, $x \neq y$. As $\{r^0, \ldots, r^l\} \cap M_f = \{r^k\}$, we have $t = \ldots f(r^i) \ldots f(r^{i+1}) \ldots$ for all $i \in \{0, \ldots, k-1\}$. As $f(r^0) = f(p) = x$ and $t = \ldots y \ldots x \ldots$ this implies that $f(r^{i+1}) \neq y$ for all $i \in \{0, \ldots, k-1\}$. Likewise $f(r^{i-1}) \neq x$ for all $i \in \{k+1, \ldots, n\}$. So, $z = f(r^k) \in A - \{x, y\}$. Now, if $t = \ldots z \ldots x \ldots$ then $\{r^0, \ldots, r^{k-1}\} \cap M_f \neq \emptyset$, a contradiction. So $t = \ldots y \ldots x \ldots z \ldots$. Likewise $\tilde{t} = \ldots x \ldots y \ldots z$. In particular if $m = 3$ then $t = yxz, \tilde{t} = xyz$. Then $\{r^0, \ldots, r^{k-1}, r^{k+1}, \ldots, r^n\} \cap M_f = \emptyset$ implies that $f(\{r^0, \ldots, r^{k-1}\}) = x$ and $f(\{r^{k+1}, \ldots, r^n\}) = y$. ∎

Note that any intermediate manipulable profile of a unanimous social choice function must contain at least two different preferences. We show that furthermore any $p \in IM_f$ that is intermediate manipulable by a coalition $S \subseteq N$ such that $p = (t^S, \tilde{t}^{N-S})$ can be manipulated by using a preference $\bar{t} \notin \{t, \tilde{t}\}$.

Lemma 41 *Let $f : P^N \to A$ be unanimous and let $(t^S, \tilde{t}^{N-S})$ be intermediate manipulable by S. Then there is a preference $\bar{t} \notin \{t, \tilde{t}\}$ such that $(f(t^S, \tilde{t}^{N-S}), f(\bar{t}^S, \tilde{t}^{N-S})) \notin t$.*

Proof. By the definition of intermediate manipulability there is a $\bar{t} \neq t$ such that $(f(t^S, \tilde{t}^{N-S}), f(\bar{t}^S, \tilde{t}^{N-S})) \notin t$. If $\bar{t} \neq \tilde{t}$ we are done, so suppose that $\bar{t} = \tilde{t}$. Then $((f(t^S, \tilde{t}^{N-S}), f(\tilde{t}^N)) \notin t$. Let $\bar{\tilde{t}} \neq \tilde{t}$ be a preference with the same top as $\tilde{t}$. Such a preference exists if $m \geq 3$. Then $f(\bar{\tilde{t}}^S, \tilde{t}^{N-S}) = f(\tilde{t}^N)$ by unanimity, and thus the coalition S can also manipulate $(t^S, \tilde{t}^{N-S})$ towards $(\bar{\tilde{t}}^S, \tilde{t}^{N-S})$, $\bar{\tilde{t}} \notin \{t, \tilde{t}\}$. ∎

5.4   Minimal manipulability of social choice functions

Let F be a nonempty set of social choice functions. For any $f \in F$ we measure its manipulability by the number of manipulable profiles $|M_f|$ that f has. We call $f^* \in F$ *minimally manipulable in F* if $|M_{f^*}| \leq |M_f|$ for all $f \in F$. Let F^* be the set of all minimally manipulable social choice functions in F. If $f^* \in F^*$ then $m_F := |M_{f^*}|$ is a lower bound on the number of manipulable profiles that any social choice function in F has.

Let F be the set of surjective nondictatorial social choice functions for fixed $m \geq 3, n \geq 2$. Then the result of Gibbard (1973) and Satterthwaite (1975) says that $m_F \geq 1$, but it does not say what the actual minimum is. This question has been solved by Kelly (1988) for two agents and three alternatives, and by Fristrup and Keiding (1998) for two agents and any number of alternatives larger than three. For reference we summarize their results in the following theorem.

Theorem 42 *Let $n = 2$. Then*

$$m_F = \begin{cases} 2 & \text{if } m = 3, \\ \frac{m!}{2} & \text{if } m \geq 4. \end{cases}$$

For the general case of $n \geq 2$, Fristrup and Keiding (1998) conjecture that $m_F = (n-1)(\frac{m!}{2} - 1) + 1$ if $(n, m) \neq (2, 3)$. We describe a class of functions that attain this number if $(n, m) \neq (2, 3)$. A social choice function f is called *almost dictatorial* if there is a profile $\bar{p} \in P^N$, an alternative $x \in A$, and an agent $d \in N$ such that $(x, dict_d(\bar{p})) \in \bar{p}(i)$ for all $i \in N - \{d\}$, and

$$f(p) = \begin{cases} dict_d(p) & \text{if } p \neq \bar{p} \\ x & \text{if } p = \bar{p}. \end{cases} \tag{5.1}$$

Proposition 43 *Let f be an almost dictatorial social choice function and let $\bar{p}, x$ be such that equation 5.1 holds. Then*

$$M_f = \{(\bar{p}_{-i}, t) \mid i \in N - \{d\}, (x, dict_d(\bar{p})) \in t, t \neq \bar{p}(i)\} \cup \{\bar{p}\}$$

and thus $|M_f| = (n-1)\left(\frac{m!}{2} - 1\right) + 1$.

Proof. Suppose that p and q are i-deviations such that $(f(p), f(q)) \notin p(i)$. This implies that if $i = d$ then $p = \bar{p}$ and if $i \neq d$ then $q = \bar{p}$. We show first that $\bar{p}$ is manipulable and treat then the case where $q = \bar{p}$. If agent d deviates to q by changing his preference in $\bar{p}$ to another preference with the same most preferred outcome we have $(f(q), f(\bar{p})) \in \bar{p}(d)$ since agent d is a dictator at q. Thus $\bar{p}$ is manipulable. Suppose that $i \in N - \{d\}$, i.e. $q = \bar{p}$. Let $p = (\bar{p}_{-i}, t), t \neq \bar{p}(i)$, be an i-deviation of q such that $(f(q), f(p)) \in p(i)$. Thus $(x, dict_d(\bar{p})) \in t$. As $(x, dict_d(\bar{p})) \in \bar{p}(i)$ there are $\frac{m!}{2} - 1$ such preferences t for every agent $i \in N - \{d\}$. This proves that $M_f = \{(\bar{p}_{-i}, t) \mid i \in N - \{d\}, (x, dict_d(\bar{p})) \in t, t \neq \bar{p}(i)\} \cup \{\bar{p}\}$.
∎

We show that the conjecture of Fristrup and Keiding cannot be true for $m = 3$. Consider the following social choice functions if $m = 3$. Let $A = \{a, b, c\}$, $t = xyz \in P$. Let $m^t : P^N \to \{a, b, c\}$ be the social choice function given by

$$m^t(p) := \begin{cases} x & \text{if } S^{(x,y)}(p) = S^{(y,z)}(p) = N, \\ y & \text{if } S^{(y,z)}(p) = N \text{ and } S^{(x,y)}(p) \neq N, \\ z & \text{if } S^{(z,y)}(p) \neq \emptyset. \end{cases}$$

Observe that m^t is surjective but not unanimous. Then we have the following proposition.

Proposition 44 *Let $t = xyz$. Then $M_{m^t} = \{(xyz^{N-\{i\}}, xzy) \mid i \in N\}$ and thus $|M_{m^t}| = n$.*

Proof. No agent will want to manipulate when $f(p) = x$. Suppose that $f(p) = y$. Then $(y, z) \in p(i)$ for all $i \in N$, so no agent can manipulate to a profile q where $f(q) = z$. Furthermore there is an agent $i \in N$ such that $(x, y) \notin p(i)$. This agent has no incentive to manipulate to the profile t^N, which is the only profile q where $f(q) = x$. On the other hand, as $p(i) \neq t$, this agent would have to change his preference in order to manipulate to the profile t^N. So f is strategy-proof at all profiles $p \in P^N$ where $f(p) \in \{x, y\}$. Suppose that $f(p) = z$. Then there is an agent $i \in N$ such that $(z, y) \in p(i)$. This agent has no incentive to manipulate to a profile q where $f(q) = y$. Again, on the other hand, as $(y, z) \in q(i)$ for such profiles, this agent would have to change his preference in order to manipulate to such a profile. The only place left where manipulations can occur is from p to t^N. If p and t^N are i-deviations we have that $p = (t^{N-\{i\}}, p(i))$. If agent i has an incentive to manipulate to t^N we must have $(x, z) \in p(i)$ and as $f(p) = z$ we must have $(z, y) \in p(i)$. Hence, $M_{m^t} = \{(t^{N-\{i\}}, xzy) \mid i \in N\}$ and $|M_{m^t}| = n$. ∎

Note that $n < (n-1)(\frac{m!}{2} - 1) + 1 = 2n - 1$ for all $n \geq 2$. Hence, the conjecture of Fristrup and Keiding cannot be true for $m = 3$. We will show however that it is true if we

replace F by the set of unanimous nondictatorial social choice functions $G \subset F$, i.e., we show that $m_G = (n-1)(\frac{m!}{2} - 1) + 1$. Moreover, we will show that then for $n \geq 3$ $g \in G^*$ if and only if g is almost dictatorial.

For the case $m = 3, n \geq 3$, we have thus $m_F \leq n < m_G = 2n - 1$. This will allow us to show that $m_F = n$ in this case, and that $f \in F^*$ for $n \geq 3$, if and only if $f = m^t$ for some $t \in P$.

5.5 Minimal manipulability with unanimity

Let $n \geq 3$. To show that $m_G = (n-1)(\frac{m!}{2} - 1) + 1$ we will use the results of Kelly (1988) and Fristrup and Keiding (1998), summarized in Theorem 42, applied to two agent social choice functions derived from $f \in G$. The first step is to embed the domain of two agent social choice functions into P^N, depending on some $S \in 2^N - \{\emptyset, N\}$. This is achieved by the map $\Pi_S : P^{\{1,2\}} \to P^N$ given by $\Pi_S(r) := (r(1)^S, r(2)^{N-S})$ for all $r \in P^{\{1,2\}}$. Now, for $f \in G$, define $f_S := f \circ \Pi_S$, then f_S is a unanimous two agent social choice function.

Clearly, any profile in $\Pi_S(M_{f_S})$ is in IM_f. Let $M_S^2 := \Pi_S(M_{f_S}) \cap M_f$. We make the following important observation, which holds by unanimity.

Lemma 45 *If $S, T \in 2^N - \{\emptyset, N\}, S \neq T$, are such that $M_S^2 \cap M_T^2 \neq \emptyset$, then $T = N - S$.*

Proof. Let $p \in M_S^2 \cap M_T^2$. Then $p = (t_1^S, \tilde{t}_1^{N-S}) = (t_2^T, \tilde{t}_2^{N-T})$ for some $t_i, \tilde{t}_i \in P$. If $T \neq N-S$ then, since $T \neq S$, this implies that $p = t^N$ for some $t \in P$. But by unanimity such p are not intermediate manipulable. Hence, $T = N - S$. ∎

Let $\mathcal{S} := \{\{1, i\} \mid i \in N\}$. Then, if $S, T \in \mathcal{S}$ are such that $S \neq T$, also $S \neq N - T$. We will show that for all $f \in G$

$$|M_f| \geq \begin{cases} 2n & \text{if } m = 3 \\ (n-1)\frac{m!}{2} & \text{if } m \geq 4 \end{cases}$$
$$> (n-1)(\frac{m!}{2} - 1) + 1,$$

if f_S is nondictatorial for all $S \in \mathcal{S}$. The simplest case is covered by the following lemma.

Lemma 46 *Suppose that all $f_S, S \in \mathcal{S}$, are nondictatorial and that $\Pi_S(M_{f_S}) \subseteq M_f$ for all $S \in \mathcal{S}$. Then*

$$|M_f| \geq \begin{cases} 2n & \text{if } m = 3, \\ \frac{m!}{2}n & \text{if } m \geq 4. \end{cases}$$

Proof. In this case, for all $S \in \mathcal{S}$, $M_S^2 = \Pi_S(M_{f_S}) \cap M_f = \Pi_S(M_{f_S})$, and thus by Theorem 42

$$|M_S^2| = |M_{f_S}| \geq \begin{cases} 2 & \text{if } m = 3, \\ \frac{m!}{2} & \text{if } m \geq 4. \end{cases}$$

By Lemma 45, $M_S^2 \cap M_T^2 = \emptyset$ for all pairwise different $S, T \in \mathcal{S}$. Since, $|\mathcal{S}| = n$ this implies the lemma. ∎

Suppose that we cannot use Lemma 46, i.e. there are $S \in \mathcal{S}$ such that $\Pi_S(M_{f_S}) - M_f \neq \emptyset$. If there is only one such S we can show the following. Note that $r \notin M_S^2$ for all $r \in P^N$ such that $|r(N)| \neq 2$.

Lemma 47 *Suppose that all $f_S, S \in \mathcal{S}$, are nondictatorial and that there is a $T \in \mathcal{S}$ such that $\Pi_S(M_{f_S}) \subseteq M_f$ for all $S \in \mathcal{S} - \{T\}$, $\Pi_T(M_{f_T}) - M_f \neq \emptyset$. Then*

$$|M_f| \geq \begin{cases} 2n & \text{if } m = 3, \\ \frac{m!}{2}(n-1) & \text{if } m \geq 4. \end{cases}$$

Proof. For $m \geq 4$ this follows in the same way as in the proof of Lemma 46. Suppose that $m = 3$. There is a $p \in \Pi_T(M_{f_T}) - M_f$. Then p is intermediate manipulable towards some $q \in P^N$, and by Lemma 41 we can assume that $|p(N) \cup q(N)| = 3$. By Corollary 39 there are at least $dist(p, q) \geq 2$ (otherwise $p \in M_f$) manipulable profiles $r \notin \{p, q\}$ on a shortest path from p to q, say r_1, r_2. As $|p(N) \cup q(N)| = 3$ and $|p(N)| = |q(N)| = 2$ we must have $|r_1(N)| = |r_2(N)| = 3$. But then $\{r_1, r_2\} \cap M_S^2 = \emptyset$ for all $S \in \mathcal{S}$. By Theorem 42, $|M_S^2| \geq 2$ for all $S \in \mathcal{S} - \{T\}$. Hence, by Lemma 45, $|M_f| \geq |\{r_1, r_2\}| + \sum_{S \in \mathcal{S}} |M_S^2| = 2n$. ∎

If there is more than one $S \in \mathcal{S}$ such that $\Pi_S(M_{f_S}) - M_f \neq \emptyset$ the manipulable profiles that we find by the intermediate manipulability of the $p \in \Pi_S(M_{f_S}) - M_f$ can be the same. The following lemma combines Lemma 41 and Corollary 39 to ensure that we are then still able to find sufficiently many manipulable different profiles $r \in P^N$ such that $|r(N)| = 3$ if there are more than three agents.

Lemma 48 *Let $S, T \in \mathcal{S}$ and let $n \geq 3$. Suppose that there are $p \in \Pi_S(M_{f_S}) - M_f$ and $\bar{p} \in \Pi_T(M_{f_T}) - M_f, p \neq \bar{p}$. Then there are $q \in \Pi_S(P^{\{1,2\}}), \bar{q} \in \Pi_T(P^{\{1,2\}})$, such that $|p(N) \cup q(N)| = |\bar{p}(N) \cup \bar{q}(N)| = 3$ and $\Pi_S^{-1}(p)$ is manipulable towards $\Pi_S^{-1}(q)$, $\Pi_S^{-1}(\bar{p})$ is manipulable towards $\Pi_S^{-1}(\bar{q})$. Suppose that there is an $r \notin \{p, \bar{p}\}$ that is on a shortest path from p to q and on a shortest path from $\bar{p}$ to $\bar{q}$. Then:*

1. *If $n \geq 4$ and $S, T \in \mathcal{S} - \{\{1\}\}$, then $S = T$, $p = \bar{q}$ and $\bar{p} = q$.*

2. *Suppose that $S = T$. Then $p = \bar{q}$ and $\bar{p} = q$. Furthermore, if $m = 3$ and there is a path between p and q that contains only one manipulable profile then p or $\bar{p}$ is intermediate manipulable by S and by $N - S$ and so $n \geq 4$. On the other hand, if $n = 3$, then all profiles on shortest paths between p and q are manipulable.*

Proof. The existence of such q and $\bar{q}$ is an immediate implication of $p \in \Pi_S(M_{f_S}) - M_f$ and $\bar{p} \in \Pi_T(M_{f_T}) - M_f$ and Lemma 41. Then $p = (t_p^S, \bar{t}_p^{N-S})$ and $\bar{p} = (t_{\bar{p}}^T, \bar{t}_{\bar{p}}^{N-T})$ and there are preferences $\bar{t}_p, \bar{t}_{\bar{p}} \in P$ such that $q \in \{(\bar{t}_p^S, \bar{t}_p^{N-S}), (t_p^S, \bar{t}_p^{N-S})\}$ and $\bar{q} \in \{(\bar{t}_{\bar{p}}^T, \bar{t}_{\bar{p}}^{N-T}), (t_{\bar{p}}^T, \bar{t}_{\bar{p}}^{N-T})\}$. By our assumptions there is an $r \notin \{p, \bar{p}\}$ that is on a shortest path from p to q and on

a shortest path from $\bar{p}$ to $\bar{q}$. Then, either $r = (t_p^S, \tilde{t}_p^{N-S-U_p}, \tilde{t}_p^{U_p})$ for some $U_p \subsetneq N - S$, $1 \le |U_p| \le |N| - |S| - 1$, or $r = (t_p^{S-V_p}, \tilde{t}_p^{V_p}, \tilde{t}_p^{N-S})$ for some $V_p \subsetneq S$, $1 \le |V_p| \le |S| - 1$. Similarly, $r = (t_{\bar{p}}^T, \tilde{t}_{\bar{p}}^{N-T-U_{\bar{p}}}, \tilde{t}_{\bar{p}}^{U_{\bar{p}}})$ for some $U_{\bar{p}} \subsetneq N - T$, $1 \le |U_{\bar{p}}| \le |N| - |T| - 1$, or $r = (t_{\bar{p}}^{T-V_{\bar{p}}}, \tilde{t}_{\bar{p}}^{V_{\bar{p}}}, \tilde{t}_{\bar{p}}^{N-T})$ for some $V_{\bar{p}} \subsetneq T$, $1 \le |V_{\bar{p}}| \le |T| - 1$. So, $\{t_p, \tilde{t}_p, \bar{t}_p\} = \{t_{\bar{p}}, \tilde{t}_{\bar{p}}, \bar{t}_{\bar{p}}\}$. As $|\{t_p, \tilde{t}_p, \bar{t}_p\}| = 3$ this implies that $S \in \{T, N - T - U_{\bar{p}}, U_{\bar{p}}\}$ or $S \in \{T - V_{\bar{p}}, V_{\bar{p}}, N - T\}$ or $N - S \in \{T, N - T - U_{\bar{p}}, U_{\bar{p}}\}$ or $N - S \in \{T - V_{\bar{p}}, V_{\bar{p}}, N - T\}$.

Proof of (1): Then $1 \in S$ and $1 \in T$ and $|S| = |T| = 2$ imply that only the first and the last case is possible, and both cases lead to $S = T$, $p = \bar{q}$ and $\bar{p} = q$ as $n \ge 4$.

Proof of (2): If $n \ge 4$ the first part of (2) follows by (1) if $S \in \mathcal{S} - \{\{1\}\}$. If $n \ge 4$ and $S = \{1\}$ or if $n = 3$ then $p = \bar{q}$ and $\bar{p} = q$ follows in a similar way using the extra assumption that $S = T$. Let $m = 3$. Let $x := f(p)$, $y := f(\bar{p})$ and $z := f(r)$. Then by Lemma 40, $p = (yxz^U, t^{N-U})$ and $\bar{p} = (xyz^U, t^{N-U})$ for some $U \in \{S, N - S\}$. Unanimity of f implies that $t \in \{zxy, zyx\}$. Without loss of generality $t = zyx$. Then p is intermediate manipulable towards (yxz^U, yzx^{N-U}) and towards $\bar{p}$, i.e. by S and $N - S$, and so $|S|, |N - S| \ge 2$ implying that $n \ge 4$.

On the other hand, if $n = 3$, then on any shortest path between p and q there is only one element, which is manipulable by Lemma 38. ∎

Now, we can prove the following theorem.

Theorem 49 *Suppose that for $m \ge 4$ all f_S, $S \in \mathcal{S} - \{\{1\}\}$, are nondictatorial, and for $m = 3$ all f_S, $S \in \mathcal{S}$, are nondictatorial. Then*

$$|M_f| \ge \begin{cases} 2n & \text{if } m = 3, \\ \frac{m!}{2}(n - 1) & \text{if } m \ge 4. \end{cases}$$

Proof. We distinguish two cases. If $n \ge 4$ we can use Lemma 48. This case is presented here. The case $n = 3$ is presented in an appendix because of its different and rather elementary technique, see Propositions 52 and 53.

So, let us assume that $n \ge 4$. For each $S \in \mathcal{S}$ and $p \in \Pi_S(M_{f_S}) - M_f$ let

$$Q(p) \quad := \quad \{q \in \Pi_S(P^{\{1,2\}}) \mid \Pi_S^{-1}(p) \text{ is manipulable towards } \Pi_S^{-1}(q)$$
$$\text{and } |p(N) \cup q(N)| = 3\}.$$

By Lemma 41 $Q(p) \ne \emptyset$, and by Corollary 39 there are at least $dist(p, q) \ge 2$ manipulable profiles for each $q \in Q(p)$ on shortest paths between p and q. Let $R(p, q)$ be the set of all these manipulable profiles for given $p \in \Pi_S(M_{f_S}) - M_f$, $S \in \mathcal{S}$, and $q \in Q(p)$.

For all $S \in \mathcal{S}$ let

$$M_S^3 := \bigcup_{p \in (\Pi_S(M_{f_S}) - M_f)} \bigcup_{q \in Q(p)} R(p, q).$$

Let $S, T \in \mathcal{S} - \{\{1\}\}$ and $p \in \Pi_S(M_{f_S}) - M_f$, $q \in Q(p)$, $\bar{p} \in \Pi_S(M_{f_S}) - M_f$, $\bar{q} \in Q(\bar{p})$. By part 1 of Lemma 48

$$R(p, q) \cap R(\bar{p}, \bar{q}) \ne \emptyset$$

implies that $S = T$ and $p = \bar{q}, \bar{p} = q$ if $n \geq 4$. Hence, if $S \neq T$, then

$$M_S^3 \cap M_T^3 = \emptyset,$$

and for each $S \in \mathcal{S}$ one $p \in \Pi_S(M_{f_S}) - M_f$ can be assumed to yield at least two new manipulable profiles in some $R(p, q), q \in Q(p)$, and thus

$$|M_S^3| \geq \frac{|\Pi_S(M_{f_S}) - M_f|}{2} * 2 = |\Pi_S(M_{f_S}) - M_f|,$$

for all $S \in \mathcal{S}$.

By definition $M_S^2 \cap M_T^3 = \emptyset$ for all $S, T \in \mathcal{S}$. Hence,

$$M_f \supseteq \bigcup_{S \in \mathcal{S}, S \neq \{1\}} M_S^2 \cup \bigcup_{S \in \mathcal{S}, S \neq \{1\}} M_S^3$$

and all these sets are pairwise disjoint, so

$$
\begin{aligned}
|M_f| &\geq \sum_{S \in \mathcal{S}, S \neq \{1\}} |M_S^2| + \sum_{S \in \mathcal{S}, S \neq \{1\}} |M_S^3| \\
&\geq \sum_{S \in \mathcal{S}, S \neq \{1\}} |\Pi_S(M_{f_S}) \cap M_f| + \sum_{S \in \mathcal{S}, S \neq \{1\}} |\Pi_S(M_{f_S}) - M_f| \\
&= \sum_{S \in \mathcal{S}, S \neq \{1\}} |\Pi_S(M_{f_S})| = \sum_{S \in \mathcal{S}, S \neq \{1\}} |M_{f_S}|.
\end{aligned}
$$

If $m \geq 4$ this proves the theorem, as $|M_{f_S}| \geq \frac{m!}{2}$ for all $S \in \mathcal{S}$ by Theorem 42.

So, suppose that $m = 3$. For all $S \in \mathcal{S}$ by Theorem 42 $|M_{f_S}| \geq 2$. If $|M_S^2| = |\Pi_S(M_{f_S}) \cap M_f| = 1$ then there is a $p \in \Pi_S(M_{f_S}) - M_f$ and $q \in Q(p)$ and thus

$$|M_S^3| \geq |R(p, q)| \geq 2.$$

If $|M_S^2| = 0$ then $|\Pi_S(M_{f_S}) - M_f| \geq 2$. Let $p, \bar{p} \in \Pi_S(M_{f_S}) - M_f, p \neq \bar{p}, q \in Q(p), \bar{q} \in Q(\bar{p})$. Then, either

$$R(p, q) \cap R(\bar{p}, \bar{q}) = \emptyset$$

and thus

$$|M_S^3| \geq |R(p, q)| + |R(\bar{p}, \bar{q})| \geq 2 + 2 = 4,$$

or

$$R(p, q) \cap R(\bar{p}, \bar{q}) \neq \emptyset$$

and without loss of generality part 2 of Lemma 48 yields that p is intermediate manipulable by S and $N - S$. So there are $q_S, q_{N-S} \in Q(p)$ such that p is intermediate manipulable towards q_S by S and intermediate manipulable towards q_{N-S} by $N - S$. Clearly, $R(p, q_S) \cap R(p, q_{N-S}) = \emptyset$ and thus

$$|M_S^3| \geq |R(p, q_S)| + |R(p, q_{N-S})| \geq 2 + 2 = 4.$$

Altogether, for all $S \in \mathcal{S}$ either

$$|M_S^2| = |\Pi_S(M_{f_S}) \cap M_f| = |\Pi_S(M_{f_S})| = |M_{f_S}| \geq 2,$$

or

$$|M_S^2| = 1 \text{ and } |M_S^3| \geq 2,$$

or

$$|M_S^2| = 0 \text{ and } |M_S^3| \geq 4.$$

We distinguish four cases, which show altogether that $|M_f| \geq 2n$.

Case 1: There are $U, V \in \mathcal{S} - \{\{1\}\}$, $U \neq V$, such that $|M_U^2| = |M_V^2| = 1$.
    Then

$$
\begin{aligned}
|M_f| &\geq \sum_{S \in \mathcal{S}, S \neq \{1\}} |M_S^2| + \sum_{S \in \mathcal{S}, S \neq \{1\}} |M_S^3| \\
&= \left(\sum_{S \in \mathcal{S}, S \notin \{\{1\}, U, V\}} |M_S^2| + |M_S^3| \right) + |M_U^2| + |M_U^3| + |M_V^2| + |M_V^3| \\
&\geq \left(\sum_{S \in \mathcal{S}, S \notin \{\{1\}, U, V\}} 2 \right) + 1 + 2 + 1 + 2 = (n-3)2 + 6 = 2n.
\end{aligned}
$$

Case 2: There is a $U \in \mathcal{S} - \{\{1\}\}$ such that $|M_U^2| = 0$.
    Then

$$
\begin{aligned}
|M_f| &\geq \sum_{S \in \mathcal{S}, S \neq \{1\}} |M_S^2| + \sum_{S \in \mathcal{S}, S \neq \{1\}} |M_S^3| \\
&= \left(\sum_{S \in \mathcal{S}, S \notin \{\{1\}, U\}} |M_S^2| + |M_S^3| \right) + |M_U^2| + |M_U^3| \\
&\geq \left(\sum_{S \in \mathcal{S}, S \notin \{\{1\}, U\}} 2 \right) + 0 + 4 = (n-2)2 + 4 = 2n.
\end{aligned}
$$

Case 3: There is a $U \in \mathcal{S} - \{\{1\}\}$ such that $|M_S^2| \geq 2$ for all $S \in \mathcal{S} - \{\{1\}, U\}$.
    Consider $M_{\{1\}}^2, M_{\{1\}}^3$. By definition also $M_{\{1\}}^2 \cup M_{\{1\}}^3 \subseteq M_f$.
    Case 3.1: $|M_{\{1\}}^2| + |M_{\{1\}}^3| \geq 4$.
    Then

$$
\begin{aligned}
|M_f| &\geq |M_{\{1\}}^2| + |M_{\{1\}}^3| + \left(\sum_{S \in \mathcal{S} - \{\{1\}, U\}} |M_S^2| \right) \\
&\geq 4 + \left(\sum_{S \in \mathcal{S} - \{\{1\}, U\}} 2 \right) = 4 + 2(n-2) = 2n.
\end{aligned}
$$

Note that in particular we have case 3.1 if $|M_{\{1\}}^2| = 0$.

Case 3.2: $|M^2_{\{1\}}| \geq 2.$
Then

$$|M_f| \geq |M^2_U| + |M^3_U| + \sum_{S \in \mathcal{S} - \{U\}} |M^3_S|$$

$$\geq 2 + \left(\sum_{S \in \mathcal{S} - \{U\}} 2 \right) = 2 + (n-1)2 = 2n.$$

Case 3.3: $|M^2_{\{1\}}| = 1$ and $|M^3_{\{1\}}| = 2.$
Then

$$|M_f| \geq |M^2_{\{1\}}| + |M^3_{\{1\}}| + \left(\sum_{S \in \mathcal{S} - \{\{1\}, U\}} |M^3_S| \right) + |M^3_U - M^3_{\{1\}}| + |M^2_U|$$

$$\geq 1 + 2 + \left(\sum_{S \in \mathcal{S} - \{\{1\}, U\}} 2 \right) + |M^3_U - M^3_{\{1\}}| + |M^2_U|$$

$$= (n-1)2 + 1 + |M^3_U - M^3_{\{1\}}| + |M^2_U| \geq 2n,$$

if $|M^2_U| \neq 0$ or $M^3_U \subsetneq M^3_{\{1\}}$. But suppose that $|M^2_U| = 0$. Then $|M^3_U| \geq 4$ and so $M^3_U \subsetneq M^3_{\{1\}}$, as $|M^3_{\{1\}}| = 2$. This finishes case 3.3, thus case 3, and proves the theorem. ∎

Let $f \in G$. We use Theorem 49 to conclude that there must be a dictatorial $f_S, S \in \mathcal{S}$, if $|M_f| \leq (n-1)(\frac{m!}{2} - 1) + 1$. With the help of this f_S we can show that f must be almost dictatorial, and thus $m_G = |M_f| = (n-1)(\frac{m!}{2} - 1) + 1$.

Theorem 50 *For all $n \geq 3$, $m_G = (n-1)(\frac{m!}{2} - 1) + 1$ and $G^* = \{f \in G \mid f$ is almost dictatorial$\}$.*

Proof. We want to prove the theorem by induction. Our induction basis is given by Theorem 42 and our induction assumption is as follows. Let $n \geq 3$. Assume that it has been shown for all $k \in \{2, \ldots, n-1\}$ and k agent unanimous nondictatorial social choice functions g that

$$|M_g| \geq \begin{cases} (k-1)(\frac{m!}{2} - 1) + 1 & \text{if } (k, m) \neq (2, 3), \\ 2 & \text{if } (k, m) = (2, 3), \end{cases}$$

and that equality holds for $k \geq 3$ if and only if g is almost dictatorial.

Induction Step:

Let $f \in G$ be such that $|M_f| \leq (n-1)(\frac{m!}{2} - 1) + 1$. Under the induction assumption we show that then $|M_f| = (n-1)(\frac{m!}{2} - 1) + 1$ and that f is almost dictatorial. This proves the theorem by the induction principle.

By Theorem 49 there must be an $S \in \mathcal{S}$ such that f_S is dictatorial. If $m \geq 4$ this S has to be in $\mathcal{S} - \{\{1\}\}$. Let $D := S$ if agent 1 is the dictator in f_S, and $D := N - S$

if agent 2 is the dictator. Let $i \in N$ and $t \in P$. We define $(n-1)$-agent social choice functions $g_{i,t} : P^{N-\{i\}} \to A$ derived from f by $g_{i,t}(p) := f(p, t^{\{i\}}), p \in P^{N-\{i\}}$. Now, for all $i \in N - D \neq \emptyset$ and $t \in P$, $g_{i,t}$ is unanimous. If such a $g_{i,t}$ is nondictatorial, then by the induction assumption

$$|M_{g_{i,t}}| \geq \begin{cases} (n-2)(\frac{m!}{2} - 1) + 1 & \text{if } (n,m) \neq (3,3), \\ 2 & \text{if } (n,m) = (3,3). \end{cases}$$

Furthermore, if such a $g_{i,t}$ is dictatorial then the dictator d must be in D. Let T_i be the set of $t \in P$ for which $g_{i,t}$ is dictatorial. Now, if $p \in M_{g_{i,t}}$ then $(p, t^{\{i\}}) \in M_f$ and clearly these profiles are different for different t. Hence, as $|M_f| \leq (n-1)(\frac{m!}{2} - 1) + 1$,

$$|T_i| \geq \begin{cases} |P| - 1 & \text{if } (n,m) \neq (3,3), \\ |P| - 2 & \text{if } (n,m) = (3,3), \end{cases}$$

as each $g_{i,t}, t \in P - T_i$, is nondictatorial and yields $|M_{g_{i,t}}|$ different manipulable profiles. We show that there is a $d \in D$ that is the dictator for all $g_{i,t}, t \in T_i$. Suppose to the contrary that there are $d_1, d_2 \in D, d_1 \neq d_2$, and $t_1, t_2 \in T_i$ such that g_{i,t_1} is dictatorial with dictator d_1 and g_{i,t_2} is dictatorial with dictator d_2. Then, agent i can choose a dictator to his advantage. Let p be a profile such that $(dict_{d_1}(p), dict_{d_2}(p)) \in p(i) = t_2$ or $(dict_{d_1}(p), dict_{d_2}(p)) \in p(i) = t_1$. Each such profile is manipulable by agent i, and there are

$$2\frac{m(m-1)}{2}(m-1)!^2(m!)^{n-3} = \left(2(m-1)(m-1)!(m!)^{n-3}\right)\frac{m!}{2} > n\frac{m!}{2} \ (n,m \geq 3)$$

such profiles, $(m-1)!^2(m!)^{n-3}$ ones for every of the $\frac{m(m-1)}{2}$ pairs $(x,y) \in p(i), x \neq y$, and a given $p(i) \in \{t_1, t_2\}, |\{t_1, t_2\}| = 2$. This contradicts $|M_f| \leq (n-1)(\frac{m!}{2} - 1) + 1$. Hence, every $g_{i,t}, t \in T_i$, is dictatorial with the same dictator $d \in D$.

As f is nondictatorial, there must be a $\bar{p} \in P^N$ such that $f(\bar{p}) \neq dict_d(\bar{p})$. Then, for all $t \in T_i$ such that $(dict_d(\bar{p}_{-i}, t), f(\bar{p})) \notin t$, $(\bar{p}_{-i}, t)$ is manipulable. For all $p \in P^N$ let

$$U(p) := \{t \in T_i \mid (dict_d(p_{-i}, t), f(p)) \notin t\},$$

then because of reflexivity of t $f(p) \neq dict_d(p_{-i}, t)$ and $(p_{-i}, t) \in M_f$ for all $t \in U(p)$. Furthermore, $|U(p)| \geq \frac{m!}{2} - |P - T_i| = |T_i| - \frac{m!}{2}$ for all p such that $f(p) \neq dict_d(p)$ and the manipulable profiles $(p_{-i}, t), (q_{-i}, t)$ that we find for such $p, q \in P^N$ are different if $p_{-i} \neq q_{-i}$. Hence,

$$(n-1)(\frac{m!}{2} - 1) + 1 \geq |M_f| \geq \sum_{t \in P-T_i} |M_{g_{i,t}}| + u\left(|T_i| - \frac{m!}{2}\right),$$

where u is the number of p having different p_{-i} and satisfying $f(p) \neq dict_d(p)$. If $(n,m) = (3,3)$ this implies $5 \geq 2 * (6 - |T_i|) + u(|T_i| - 3)$ which, as $|T_i| \in \{4, 5\}$, shows that $u = 1$. On the other hand, if $(n,m) \neq (3,3)$ we use $|T_i| = m! - 1$ and obtain $(n-1)(\frac{m!}{2} - 1) + 1 \geq (n-2)(\frac{m!}{2} - 1) + 1 + u(\frac{m!}{2} - 1)$, which also implies that $u = 1$.

Now, note that $u = 1$ implies that $f(p) = dict_d(p)$ for all $p \in P^N$ such that $p_{-i} \neq \bar{p}_{-i}$. We show that $g_{i,\bar{p}(i)}$ is almost dictatorial. Clearly, $\bar{p}$ is manipulable by agent d. Let $j \in N - \{i, d\}$. Then all $(\bar{p}_{-j}, t)$ such that $(dict_d(\bar{p}), f(\bar{p})) \notin t \neq \bar{p}(j)$ are manipulable. This yields

$$|M_{g_{i,\bar{p}(i)}}| \geq |N - \{i, d\}|(\frac{m!}{2} - 1) + 1 = (n - 2)(\frac{m!}{2} - 1) + 1,$$

where equality holds if and only if $(f(\bar{p}), dict_d(\bar{p})) \in \bar{p}(j)$ for all $j \in N - \{i, d\}$. Hence, $g_{i,\bar{p}(i)}$ is almost dictatorial. If $(n, m) = (3, 3)$ this implies that $|T_i| = m! - 1$ as otherwise

$$5 \geq |M_f| \geq \sum_{t \in P - T_i} |M_{g_{i,t}}| \geq 3 + 3 = 6,$$

a contradiction.

Finally,

$$(n - 1)(\frac{m!}{2} - 1) + 1 \geq |M_f| \geq |M_{g_{i,\bar{p}(i)}}| + |U(\bar{p})| \geq (n - 2)(\frac{m!}{2} - 1) + 1$$

shows that $|U(\bar{p})| = (\frac{m!}{2} - 1)$ and so $(f(\bar{p}), dict_d(\bar{p})) \in \bar{p}(i)$ and f is almost dictatorial. ∎

5.6   Minimal manipulability with three alternatives and surjectivity

Theorem 51 *Let $n \geq 3$, $m = 3$ and let F be the set of surjective nondictatorial social choice functions. Then $m_F = n$ and $F^* = \{m^t \mid t \in P\}$.*

Proof. Let $f \in F^*$. Since m^t is surjective and nondictatorial we have $|M_f| \leq |M_{m^t}| = n$. If f is unanimous then by Theorem 50 $|M_f| \geq 2(n - 1) + 1 > n$, a contradiction. So f is not unanimous and there are $x \in A = \{a, b, c\}$ and $\tilde{p} \in top(x)$ such that $f(\tilde{p}) \neq x$. Fix such an x and $\tilde{p}$, and let $\{y, z\} = A - \{x\}$.

Claim 1: $a \in f(top(a))$ for all $a \in A$.

To the contrary suppose $a \notin f(top(a))$. We deduce that $|M_f| > n$. As f is surjective there exist $p \in P^N$ such that $f(p) = a$. Then $a \notin f(top(a))$ implies that for some $j \in N$ we have $p(j) \notin \{abc, acb\}$. Without loss of generality $j = 1$. Furthermore, let p be chosen such that for all q, with $f(q) = a$ we have $dist(q, top(a)) \geq dist(p, top(a))$. For $i \geq 2$ let $\bar{r}^i := (t, p_{-i})$, where $t \in \{abc, acb\} - \{p(i)\}$ and let $\bar{\bar{r}}^i := (abc, \bar{r}^i_{-1})$. Furthermore let $\bar{r}^1 := (abc, p_{-1})$ and $\hat{r}^1 := (acb, p_{-1})$. As for $i \geq 2$

$$dist(\bar{\bar{r}}^i, top(a)) < dist(\bar{r}^i, top(a)) \leq dist(p, top(a))$$

and

$$dist(\bar{r}^1, top(a)) = dist(\hat{r}^1, top(a)) < dist(p, top(a))$$

we have $f(\bar{\bar{r}}^i) \neq a$ for $i \geq 2$, $f(\bar{r}^1) \neq a$ and $f(\hat{r}^1) \neq a$. So $\bar{r}^1, \hat{r}^1 \in M_f$ and $M_f \cap \{\bar{r}^i, \bar{\bar{r}}^i\} \neq \emptyset$ for all $i \geq 2$. Hence $|M_f| \geq n + 1$, the desired contradiction, and claim 1 is proven.

Let $V := top(x) \cap M_f$. Claim 1 shows that there exist $q \in top(x)$ such that $f(q) = x$. Now consider a path from $\hat{p}$ to q through $top(x)$. On this path at least one profile is manipulable. So, $V \neq \emptyset$. Let $W := \{p \in top(x) \mid f(p) \neq x\}$. Obviously, $V \subseteq W$.

Claim 2: $|W| > n$.

To the contrary suppose $|W| \leq n$. Consider $\bar{f} : P^N \to A$ defined by

$$\bar{f}(q) := \begin{cases} x & q \in W \\ f(q) & \text{otherwise.} \end{cases}$$

We show that $|M_{\bar{f}}| < |M_f|$ and have a contradiction with minimal manipulability of f if $\bar{f}$ is surjective and nondictatorial.

- $\bar{f}$ is surjective By claim 1 $y \in f(top(y))$ for all $y \in A$. So $\bar{f}$ is surjective.

- $\bar{f}$ is nondictatorial Suppose to the contrary that $\bar{f}$ is dictatorial with dictator d. There is a $\hat{p} \in W$ with $f(\hat{p}) \neq x$. By the definition of $\bar{f}$ we have $f(p) = \bar{f}(p) = dict_d(p)$ for all $p \notin W$. For $i \in N - \{d\}$ consider i-deviations q of $\hat{p}$ such that $(x, f(\hat{p})) \notin q(i)$. Then f is manipulable at q by i towards $\hat{p}$, as $q \notin W$ and thus $f(q) = dict_d(q) = dict_d(\hat{p}) = x$. As there are three preferences t in P at which $(x, f(\hat{p}), x) \notin t$ there are at least $2(n-1)$ of these q-profiles. Since, $3 \leq n$ we obtain $|M_f| \geq 2(n-1) > n$, a contradiction with the assumption that $|M_f| \leq n$. Thus f is nondictatorial.

- $|M_{\bar{f}}| < |M_f|$ By the assumption $|W| \leq n$, there is for all $q \in W$ an adjacent $r \in top(x)$, i.e. $dist(r, q) = 1$, such that $f(r) = x$. So f is manipulable at all $q \in W$. Therefore, $V = W$. We show that the transition from f to $\bar{f}$ repairs more manipulable profiles than it creates. All the manipulable profiles in V are not manipulable in $\bar{f}$ anymore. Let $q \in M_{\bar{f}} - M_f$ be a manipulable profile that was created by the transition. Then there is an agent i and a profile $p \in W$ such that p and q are i-deviations satisfying $(f(q), x) \notin q(i) \notin \{xyz, xzy\}$. This leads to the two possibilities $f(q) = z$ and $q(i) = yxz$ or $f(q) = y$ and $q(i) = zxy$. Then $q \notin M_f$ implies in the first case that $f(r) = z \neq x$ and in the second case that $f(r) = y \neq x$ for all i-deviations r from q. If $r(i) \neq p(i)$, r is also an i-deviation from p. Hence, for $t \in \{xyz, xzy\} - \{p(i)\}$ we have $f(p_{-i}, t) \neq x$. As $(p_{-i}, t) \in V = W, (p_{-i}, t) \neq p$, there are at most $(|V| - 1)$ agents i such that there is a $q \in M_{\bar{f}} - M_f$ manipulable by i towards a profile p in W. Let m_i be the number of such profiles q for agent i. As $q(i) = yxz$ if $f(q) = z$ and $q(i) = zxy$ if $f(q) = y$, we have $m_i \leq 1$. So, $|M_{\bar{f}}| \leq |M_f| - |V| + (|V| - 1)m_i \leq |M_f| - 1$. This proves claim 2.

Claim 3: $|V| = n$.

By claim 2 $|W| \geq n + 1$. By claim 1 $top(x) - W \neq \emptyset$. So, there are $p, q \in P^N$ such that $p \in W$ and $q \in top(x) - W$. Hence, $f(p) \neq x$ and $f(q) = x$. Without loss of generality let $p \in W - V$, if such p does not exist we have $|M_f| \geq |V| \geq n + 1$, a contradiction. As

$p \notin M_f$ all profiles $r^1, r^2, \ldots, r^n$ adjacent to p in $top(x)$ are in W. Consider a path from r^j to q, say $r^j_1, \ldots, r^j_{k_j} \in top(x)$. Then $V \cap \{r^j_1, \ldots, r^j_{k_j-1}\} \neq \emptyset$. So, if we can find disjoint paths for every j from r^j to q we have $|V| \geq n$. We show how to construct such paths. Without loss of generality there is a $k \in \mathbb{N}$ such that

$$p(i) = q(i) \text{ for } n \geq i > k,$$
$$p(i) \neq q(i) \text{ for } 1 \leq i \leq k,$$

and the $r^1, r^2, \ldots, r^n$ are such that $r^j = (p_{-i}, t), t \in \{xyz, xzy\} - \{p(j)\}$ for all $j \in \{1, 2, \ldots, n\}$. Hence, $r^j(i) = p(i)$ if $i \neq j$, $r^j(j) = q(j)$ if $j \leq k$, and $r^j(j) \neq q(j) = p(j)$ if $j > k$. Let $1 \leq j_1 \leq n, 1 \leq j_2 < k$. For $j_1 \leq k, j_2 < k$, let

$$r^{j_1}_{j_2}(i) := \begin{cases} q(i) & \text{if } j_1 < i \leq j_1 + j_2 \leq k \text{ or} \\ & i \leq j_1 + j_2 - k \text{ and } j_1 + j_2 > k \\ p(i) & \text{otherwise,} \end{cases}$$

and for $j_1 > k, j_2 \leq k$, let

$$r^{j_1}_{j_2}(i) := \begin{cases} q(i) & \text{if } i \leq j_2 \\ p(i) & \text{otherwise.} \end{cases}$$

Then these are disjoint paths $r^{j_1}, 1 \leq j_1 \leq n$, between p and q. So, $|V| \geq n$ and since $|V| \leq |M_f| \leq n$ this proves claim 3.

Now, $n = |V| = |M_f \cap top(x)| \leq |M_f| \leq n$. So, $M_f \subseteq top(x)$. By claim 3 $|W| \geq n+1$. So, there is a profile $p \in W - V$.

<u>Claim 4:</u> $f^{-1}(\{x\}) \subseteq top(x)$.
Suppose contrapositive without loss of generality that there is an $r \in P^N$ such that $f(r) = x$ and $r(n) \notin \{xyz, xzy\}$. Then

$$r^k(i) = \begin{cases} r(i) & \text{if } k < i \\ p(i) & \text{otherwise} \end{cases}$$

defines a path from $r = r^0$ to $r^n = p$ outside $top(x)$, i.e. $r^k \notin top(x)$ for all $k < n$. Since $M_f \cap \{r^0, r^1, \ldots, r^{n-1}\} \neq \emptyset$, we have a contradiction because $M_f \subseteq top(x)$. This proves claim 4.

Let $U := P^N - top(x)$. Then $U \cap M_f = \emptyset$ and $f(U) = \{y, z\}$.

<u>Claim 5:</u> For profiles $p \in P^N$ and $q \in U$ such that $f(p) = y$ and $S^{(y,z)}(p) \subseteq S^{(y,z)}(q) \neq \emptyset$, we have $f(q) = y$.
By claim 4 $x \notin f(U)$, so $f(q) \neq x$. Without loss of generality p, q are i-deviations and $q(i) \in \{yxz, yzx\}$. As $q \notin M_f$, $f(q) = y$. This proves claim 5.

Now fix $y, z \in A - \{x\}$. Let

$$W^y \; : \; = \{S \subseteq N \mid \text{for all } p \in P^N : f(p) \neq x \text{ and } S^{(y,z)}(p) = S$$
$$\text{implies } f(p) = y\},$$
$$W^z \; : \; = \{S \subseteq N \mid \text{for all } p \in P^N : f(p) \neq x \text{ and } S^{(z,y)}(p) = S$$
$$\text{implies } f(p) = z\}.$$

Then $N \in W^y$ by surjectivity and claim 5, as $S^{(y,z)}(p) \subseteq N = S^{(y,z)}(q)$ for all p such that $f(p) = y$ and $q \in top(y) \subset U$. Likewise, $N \in W^z$.

Claim 6: $f(xyz^S, xzy^{N-S}) = x$ implies $S = N$ or $S = \emptyset$.

To the contrary let $i \in S$ and $j \in N - S$. Consider $r = (yxz^{\{i\}}, xyz^{S-\{i\}}, xzy^{N-S})$. As $r \in U$, and therefore $r \notin M_f$, we have $f(r) \neq z$. But similarly $f(\bar{r}) \neq y$, where $\bar{r} = (zxy^{\{j\}}, xyz^S, xzy^{N-S-\{j\}})$. By claim 4 $f(r) \neq x$. Thus $f(r) = y$ and $f(\bar{r}) \neq y, \bar{r} \in U$. As $S^{(y,z)}(r) = S^{(y,z)}(\bar{r}) \neq \emptyset$ we have a contradiction with claim 5. So $S = N$ or $S = \emptyset$, proving claim 6.

But then $W^y = 2^N \backslash \{\emptyset\}$ and $W^z = \{N\}$ or $W^y = \{N\}$ and $W^z = 2^N \backslash \{\emptyset\}$. Moreover, we have either $f^{-1}(\{x\}) = \{xyz^N\}$ if $W^z = 2^N \backslash \{\emptyset\}$, or $f^{-1}(\{x\}) = \{xzy^N\}$ if $W^y = 2^N \backslash \{\emptyset\}$. So, $f = m^t$ for some $t \in \{xyz, xzy\}$. ∎

5.7 Conclusion

In this chapter we have found the minimally manipulable surjective and nondictatorial social choice functions with three alternatives and more than two agents. They turn out to be anonymous. This is in contrast to the social choice functions attaining the global minimum 0 in the class of all social choice functions, which are the dictatorial ones. The second smallest value that the function $K(f) := |M_f|$, f a social choice function for three alternatives, takes is n. This value is not attained, as one might expect, by the almost dictatorial social choice functions that give the minimum in the case of two agents and four or more alternatives, see Fristrup and Keiding (1998), but it is attained for all numbers of agents larger than two by six anonymous social choice functions similar in structure. For the case of two agents these six social choice functions are only part of the set of all minimally manipulable social choice functions, see Kelly (1988) for other examples of social choice functions attaining the minimum in this case. For more than three alternatives it is not clear whether the almost dictatorial social choice functions are minimally manipulable for more than two agents. We show this however if surjectivity is replaced by unanimity. In that case almost dictatorial social choice functions are the minimally manipulable social choice functions if in addition $n \geq 3$. The case of four alternatives and three agents is after this chapter the smallest unsolved case with surjectivity, and for two agents the minimally manipulable rules are not characterized. It would be of interest to see whether only almost dictatorial social choice functions are best in the case of two agents and more than three

alternatives, as this might help with an induction. The following table summarizes these results about minimal manipulability. FK (1998) stands for Fristrup and Keiding (1998), and K (1988), K (1989) stand for Kelly (1988) and Kelly (1989) respectively.

	$m = 3$		
n	m_F	$f \in F^*$	in:
$= 2$	2	if $f = m^t, t \in P$	K (1988) m_F
≥ 3	n	iff $f = m^t, t \in P$	Section 5.6
	m_G	$g \in G^*$	
$= 2$	2	examples	K (1989)
≥ 3	$2n - 1$	iff almost dictatorial	Section 5.5
	$m \geq 4$		
n	m_F	$f \in F^*$	
$= 2$	$\frac{m!}{2}$	if almost dictatorial	FK (1998)
≥ 3	$\leq (n-1)(\frac{m!}{2} - 1) + 1$	?	FK (1998)
	m_G	$g \in G^*$	
$= 2$	$\frac{m!}{2}$	if almost dictatorial	FK (1998)
≥ 3	$(n-1)(\frac{m!}{2} - 1) + 1$	iff almost dictatorial	Section 5.5

We relate the result in Section 5.6 also to a conjecture made in Kelly (1988). The Kemeny distance between two preferences t and t' is defined to be the minimal number of transpositions of adjacent positions in the preference t necessary to obtain the preference t'. Local strategy-proofness is said to hold if there is a $\delta > 1$ such that to manipulate, an agent always has to change to a preference t' that has a larger Kemeny distance from his true preference t than δ. In the minimally manipulable social choice functions characterized here, the manipulating agent changes from xyz to xzy. So the Kemeny distance is 1. Now, in Kelly (1988) the conjecture was that local strategy-proof social choice functions and minimally manipulable social choice functions have a nonempty intersection. So instead of looking for minimally manipulable social choice functions one can look for local strategy-proof social choice functions, which might be easier. The result from Section 5.6 says that this is not true for three alternatives and more than two agents. This relation is also not true for almost dictatorial social choice functions and two agents. But since Fristrup and Keiding (1998) do not prove that these are all minimally manipulable social choice functions for two agents, this does not say that the relation is not true in this case. So, it is of interest not only to find the minimum in the other cases than the ones considered here, but to characterize also all social choice functions that give this minimum, as was achieved here. In the case of three alternatives and two agents the conjecture will remain true of course, since this was the example given in Kelly (1988).

In another interesting paper by Kelly (1993), a computer draws social choice functions uniformly from all social choice functions satisfying axioms like anonymity, neutrality and Pareto optimality, or combinations thereof. He then investigates the sample distributions

with respect to the number of manipulable profiles of these social choice functions. One sees that imposing anonymity in the case of three alternatives causes a shift towards social choice functions with less manipulable profiles, compared to the sample obtained without any constraints. We saw here that the minimally manipulable social choice functions for three alternatives are anonymous. So in this sense anonymity is for three alternatives a good property when looking for minimally manipulable social choice functions. It would be interesting to see what the distributions look like for more than three alternatives, to see what the effect is there.

5.8   Appendix

In this appendix we study the case where $n = 3$ and so Lemma 48 cannot be used to show that $M_S^3 \cap M_T^3 = \varnothing$ if $S, T \in \mathcal{S}$, $S \neq T$. Instead we use even more elementary techniques.

Proposition 52 *Let* $f : P^{\{1,2,3\}} \rightarrow \{x, y, z\}$ *be a unanimous and nondictatorial social choice function. Suppose that all $f_S, S \in \mathcal{S}$, are nondictatorial. Then*

$$|M_f| \geq 6.$$

Proof. By Theorem 42 there are $p_1, p_2 \in M_{f_{\{1,2\}}}, p_3, p_4 \in M_{f_{\{1,3\}}}$ and $p_5, p_6 \in M_{f_{\{2,3\}}}$ such that

$$p_1 = (t_1, t_1, \tilde{t}_1), p_3 = (t_3, \tilde{t}_3, t_3), p_5 = (\tilde{t}_5, t_5, t_5),$$
$$p_2 = (t_2, t_2, \tilde{t}_2), p_4 = (t_4, \tilde{t}_4, t_4), p_6 = (\tilde{t}_6, t_6, t_6),$$

and $p_i \neq p_{i+1}, i \in \{1, 3, 5\}$, $t_i \neq \tilde{t}_i, i \in \{1, 2, 3, 4, 5, 6\}$, as otherwise these profiles cannot be manipulable because of unanimity. This also implies that $|\{p_1, p_2, p_3, p_4, p_5, p_6\}| = 6$, since otherwise $t_i = \tilde{t}_i$ for some $i \in \{1, 2, 3, 4, 5, 6\}$. For each p_i there is a $\bar{t}_i \in P$ and an $S \in \mathcal{S}$, such that $p_i \notin M_f$ and p_i is manipulable towards $q_i := (\bar{t}_i^S, \tilde{t}_i)$ by S or $p_i \in M_f$ and p_i is manipulable towards $q_i := (t_i^S, \bar{t}_i)$ by $N - S$. If $p_i \notin M_f$ then $\bar{t}_i \neq t_i$ and if $p_i \in M_f$ then $\bar{t}_i \neq \tilde{t}_i$.

Claim: Let $p_i \notin M_f$. Then $|M_f| \geq 6$ or p_i is manipulable towards some $q_i = (\bar{t}_i^S, \tilde{t}_i), S \in \mathcal{S}$, such that $\bar{t}_i \notin \{t_i, \tilde{t}_i\}$.

As $p_i \notin M_f$ a $q_i = (\bar{t}_i^S, \tilde{t}_i)$ towards which p_i is manipulable must exist and as $p_i \neq q_i$ $t_i \neq \bar{t}_i$. Suppose to the contrary that $|M_f| \leq 5$ and $\bar{t}_i \in \{t_i, \tilde{t}_i\}$, i.e. $\bar{t}_i = \tilde{t}_i$. Without loss of generality let $S = \{1, 2\}$, i.e. $i \in \{1, 2\}$. We omit the subscript i. Let $u := f(p), v := f(q)$. Then $v \neq u$ and $(v, u) \in t$. As $\bar{t} = \tilde{t}$ and $f(q) = f(\tilde{t}, \tilde{t}, \tilde{t}) = v$ unanimity implies that $\tilde{t} = \bar{t} \in \{vwu, vuw\}$. Then, $t = wvu$ as by unanimity and $f(p) \neq v$ $t \in P - \{vwu, vuw\}$, and $(v, u) \in t$. This in turn implies that $\tilde{t} = \bar{t} = vuw$ as otherwise $p = (wvu, wvu, vwu)$ is manipulable under f by $\{3\}$ towards (wvu, wvu, wvu) by unanimity. Hence,

$$p = (wvu, wvu, vuw), q = (vuw, vuw, vuw).$$

Now, unanimity implies that p is also manipulable towards (vwu, vwu, vuw) by the coalition $\{1, 2\}$ and so, as $p \notin M_f$,

$$M_f \supseteq \{(t, \bar{t}, \bar{t}), (\bar{t}, t, \bar{t}), (t, vwu, \bar{t}), (vwu, t, \bar{t})\}.$$

Consider p_3, p_4, p_5 and p_6. Then $|M_f| \leq 5$ implies that $|M_f \cap \{p_3, p_4, p_5, p_6\}| \leq 1$. So, without loss of generality $p_3, p_4 \notin M_f$. But then

$$M_f \supseteq \{(t_3, \tilde{t}_3, \bar{t}_3), (\bar{t}_3, \tilde{t}_3, t_3), (t_4, \tilde{t}_4, \bar{t}_4), (\bar{t}_4, \tilde{t}_4, t_4)\},$$

and $|M_f| \leq 5$ implies that without loss of generality $\{(t_3, \tilde{t}_3, \bar{t}_3), (\bar{t}_3, \tilde{t}_3, t_3)\} = \{(\bar{t}, t, \bar{t}), (vwu, t, \bar{t})\}$, since $(\bar{t}, t, \bar{t}), (vwu, t, \bar{t})$ are the only profiles in

$$\{(t, \bar{t}, \bar{t}), (\bar{t}, t, \bar{t}), (t, vwu, \bar{t}), (vwu, t, \bar{t})\}$$

where the second agent has an identical preference. But this implies that $t_3 = \bar{t} = \bar{t}_3$, a contradiction. This finishes the proof of the claim.

Now, we consider three cases.

Case 1: There is an $i \in \{1, 3, 5\}$ such that $p_i, p_{i+1} \notin M_f$.

Without loss of generality $i = 1$. By the claim there are $\bar{t}_1, \bar{t}_2 \in P$ such that $|\{t_1, \tilde{t}_1, \bar{t}_1\}| = |\{t_2, \tilde{t}_2, \bar{t}_2\}| = 3$ and

$$M_f \supseteq \{(t_1, \bar{t}_1, \tilde{t}_1), (\bar{t}_1, t_1, \tilde{t}_1), (t_2, \bar{t}_2, \tilde{t}_2), (\bar{t}_2, t_2, \tilde{t}_2)\}.$$

We show that $|\{(t_1, \bar{t}_1, \tilde{t}_1), (\bar{t}_1, t_1, \tilde{t}_1), (t_2, \bar{t}_2, \tilde{t}_2), (\bar{t}_2, t_2, \tilde{t}_2)\}| = 4$. So, suppose contrapositive that $|\{(t_1, \bar{t}_1, \tilde{t}_1), (\bar{t}_1, t_1, \tilde{t}_1), (t_2, \bar{t}_2, \tilde{t}_2), (\bar{t}_2, t_2, \tilde{t}_2)\}| \leq 3$. As

$$|\{t_1, \bar{t}_1, \tilde{t}_1\}| = |\{t_2, \tilde{t}_2, \bar{t}_2\}| = 3$$

and $p_1 \neq p_2$ this implies that $t_1 = \bar{t}_2, \bar{t}_1 = t_2$ and $\tilde{t}_1 = \tilde{t}_2$. Hence, if we let $t = t_1, \bar{t}_1 = \bar{t}$ and $\tilde{t}_1 = \tilde{t}$ then $p_1 = (t, t, \bar{t}) = q_2$ and $p_2 = (\bar{t}, \bar{t}, \tilde{t}) = q_1$. An application of Lemma 40 yields that $t = uvw$ and $\bar{t} = vuw$ where $v = f(p_1)$ and $u = f(p_2)$. By unanimity then $\tilde{t} \in \{wuv, wvu\}$. But then, if $\tilde{t} = wuv$ $p_1 \in M_f$ by unanimity and if $\tilde{t} = wvu$ $p_2 \in M_f$ by unanimity, a contradiction. Hence,

$$|\{(t_1, \bar{t}_1, \tilde{t}_1), (\bar{t}_1, t_1, \tilde{t}_1), (t_2, \bar{t}_2, \tilde{t}_2), (\bar{t}_2, t_2, \tilde{t}_2)\}| = 4.$$

Note that we are now in a similar situation as in the end of the proof of the claim, except that we have additionally that $|\{t_1, \bar{t}_1, \tilde{t}_1\}| = 3$ and $|\{t_2, \tilde{t}_2, \bar{t}_2\}| = 3$. So, by considering p_3, p_4, p_5 and p_6, we can obtain the same contradiction. This finishes case 1.

Case 2: There is an $i \in \{1, 3, 5\}$ such that $p_i, p_{i+1} \in M_f$.

Then $|M_f| \leq 5$ and case 1 implies that without loss of generality either $p_1, p_2, p_3, p_5 \in M_f$ and $p_4, p_6 \notin M_f$ or $p_1, p_2, p_3, p_4, p_5 \in M_f$ and $p_6 \notin M_f$. In both cases we can choose $t_6, \bar{t}_6, \tilde{t}_6 \in P$ by the claim such that $|\{t_6, \bar{t}_6, \tilde{t}_6\}| = 3$ and $(\tilde{t}_6, t_6, \bar{t}_6), (\bar{t}_6, \tilde{t}_6, t_6) \in M_f$. Then

$$\{p_1, p_2, p_3, p_5\} \cap \{(\tilde{t}_6, t_6, \bar{t}_6), (\bar{t}_6, \tilde{t}_6, t_6)\} = \varnothing,$$

contradicting $|M_f| \leq 5$. This finishes case 2.

Case 3: For all $i \in \{1,3,5\}$ either $p_i \in M_f$ and $p_{i+1} \notin M_f$ or $p_i \notin M_f$ and $p_{i+1} \in M_f$.

Without loss of generality $p_1, p_3, p_5 \in M_f$. By the claim there are $t_i, \bar{t}_i, \tilde{t}_i \in P, i \in \{2,4\}$, such that

$$M_f \supseteq \{(t_2, \bar{t}_2, \tilde{t}_2), (\bar{t}_2, t_2, \tilde{t}_2), (t_4, \bar{t}_4, \tilde{t}_4), (\bar{t}_4, \tilde{t}_4, t_4)\}$$

and $|\{t_i, \bar{t}_i, \tilde{t}_i\}| = 3$ for all $i \in \{2,4\}$. Then $|M_f| \leq 5$ implies that

$$\{(t_2, \bar{t}_2, \tilde{t}_2), (\bar{t}_2, t_2, \tilde{t}_2)\} = \{(t_4, \tilde{t}_4, \tilde{t}_4), (\bar{t}_4, \tilde{t}_4, t_4)\},$$

and so $\bar{t}_2 = \tilde{t}_4 = t_2$, a contradiction. This finishes case 3.

All three cases together show that under the assumptions of the proposition $|M_f| \leq 5$ leads to a contradiction. Hence, $|M_f| \geq 6$, proving the proposition. ∎

Proposition 53 *Let $f : P^{\{1,2,3\}} \to A$ be a unanimous and nondictatorial social choice function. Suppose that all $f_S, S \in \mathcal{S} - \{\{1\}\}$, are nondictatorial. Then*

$$|M_f| \geq m! > 2\left(\frac{m!}{2} - 1\right) + 1.$$

Proof. As $N = \{1,2,3\}$ we have $\mathcal{S} - \{\{1\}\} = \{\{1,2\}, \{1,3\}\}$. By Lemma 42 $|\Pi_S(M_{f_S})| \geq \frac{m!}{2}$ for all $S \in \mathcal{S} - \{\{1\}\}$. Without loss of generality $k := |\Pi_{\{1,2\}}(M_{f_{(1,2)}}) \cap M_f| \leq |\Pi_{\{1,3\}}(M_{f_{(1,3)}}) \cap M_f|$. If the following claim is proven the proposition follows immediately, as then

$$\begin{aligned}
|M_f| &\geq 2 * |\Pi_{\{1,2\}}(M_{f_{(1,2)}}) - M_f| + |\Pi_{\{1,2\}}(M_{f_{(1,2)}}) \cap M_f| \\
&\quad + |\Pi_{\{1,3\}}(M_{f_{(1,3)}}) \cap M_f| \\
&\geq 2 * \left(\frac{m!}{2} - k\right) + k + k = m! > 2\left(\frac{m!}{2} - 1\right) + 1.
\end{aligned}$$

Claim: To each $p \in \Pi_{\{1,2\}}(M_{f_{(1,2)}}) - M_f$ we can associate two profiles $r_p^1, r_p^2 \in M_f$ such that $|r_p^1(N)| = |r_p^2(N)| = 3$ and $\{r_p^1, r_p^2\} \cap \{r_{\tilde{p}}^1, r_{\tilde{p}}^2\} = \emptyset$ for all $\tilde{p} \in \Pi_{\{1,2\}}(M_{f_{(1,2)}}) - M_f, \tilde{p} \neq p$.

Let $p = (t_p, t_p, \tilde{t}_p) \notin M_f$ be manipulable towards $q = (\bar{t}_p, \bar{t}_p, \tilde{t}_p)$. By Lemma 41 we can assume that $\bar{t}_p \notin \{t_p, \tilde{t}_p\}$. For all such p by Lemma 38 $r_p^1 := (t_p, \bar{t}_p, \tilde{t}_p) \in M_f$ and $r_p^2 := (\bar{t}_p, t_p, \tilde{t}_p) \in M_f$. If $\{r_p^1, r_p^2\} \cap \{r_{\tilde{p}}^1, r_{\tilde{p}}^2\} = \emptyset$ for all $p, \tilde{p} \in \Pi_{\{1,2\}}(M_{f_{(1,2)}}) - M_f, \tilde{p} \neq p$, we are done.

So, suppose that there is an $r \in \{r_p^1, r_p^2\} \cap \{r_{\tilde{p}}^1, r_{\tilde{p}}^2\}$ for some $p, \tilde{p} \in \Pi_{\{1,2\}}(M_{f_{(1,2)}}) - M_f$, $\tilde{p} \neq p$. By Lemma 48 then $p = (t,t,\bar{t}) = \tilde{q}$ and $q = (\bar{t}, \bar{t}, \tilde{t}) = \tilde{p}$ for some $t, \bar{t}, \tilde{t} \in P$ such that $|\{t, \bar{t}, \tilde{t}\}| = 3$. Let $x := f(p)$ and $f(\tilde{p}) = y$. Then $x \neq y$, $(y,x) \in t$ and $(x,y) \in \bar{t}$ because f is manipulable from p to $\tilde{p}$ and from $\tilde{p}$ to p. Without loss of generality we may assume that $(x,y) \in \tilde{t}$. As f is unanimous and $f(\tilde{p}) = y$ either $\tilde{t} \neq x \ldots$ or $\bar{t} \neq x \ldots$. As, $m \geq 4$ there

are $(m-2)! \geq 2$ preferences $\widehat{t}$ such that $\widehat{t} = xy\ldots$, hence we can choose such a $\widehat{t} \notin \{\widetilde{t}, \widehat{t}\}$. Let $\widehat{p} = (\widehat{t}, \widehat{t}, \widehat{t})$ and

$$\widehat{r}^1 = (\widehat{t}, \widetilde{t}, \widetilde{t}), \widehat{r}^2 = (\widetilde{t}, \widehat{t}, \widetilde{t}),$$
$$\overline{r}^1 = (\widehat{t}, t, \widetilde{t}), \overline{r}^2 = (t, \widehat{t}, \widetilde{t}).$$

We need the following subclaim concerning these profiles.

Subclaim: Let $\widehat{r}^i \notin M_f$ for some $i \in \{1,2\}$. Then $\overline{r}^i \in M_f$ and $\widehat{p} \notin \Pi_{\{1,2\}}(M_{f_{\{1,2\}}}) - M_f$. Furthermore then for all $p' \in \Pi_{\{1,2\}}(M_{f_{\{1,2\}}}) - M_f$ such that p' is manipulable by $\{1,2\}$ towards q' with $|p'(N) \cup q'(N)| = 3$ and $\widehat{r}^i$ or $\overline{r}^i, i \in \{1,2\}$, is on a shortest path from p' to q', $p' = \widetilde{p}$ or $p' = p$ respectively.

Without loss of generality $i = 1$. If $f(\widehat{r}^1) \in A - \{x,y\}$ then agent 1 can manipulate at $\widehat{r}^1$ towards $\widetilde{p}$, contradicting $\widehat{r}^1 \notin M_f$. If $f(\widehat{r}^1) = x$ then $\widetilde{p}$ is manipulable by agent 1 towards $\widehat{r}^1$, contradicting $\widetilde{p} \notin M_f$. Thus

$$f(\widehat{r}^1) = y.$$

If $f(\widehat{p}) = y$, then $\widehat{p} \in M_f$ as agent 3 can deviate towards $(\widehat{t}, \widehat{t}, \widehat{t})$ and $(x,y) \in \widetilde{t}$. If $f(\widehat{p}) \in A - \{x,y\}$ then $\widehat{p} \in M_f$ as agent 2 can deviate towards $\widehat{r}^1$ and $\widehat{p}(2) = \widehat{t} = xy\ldots$. If $f(\widehat{p}) = x$ then $\widehat{p} \notin \Pi_{\{1,2\}}(M_{f_{\{1,2\}}}) - M_f$ as $\widehat{p}(\{1,2\}) = \widehat{t} = x\ldots$, so coalition $\{1,2\}$ has no incentive to manipulate. Hence, for all values of $f(\widehat{p})$ either $\widehat{p} \in M_f$ or $\widehat{p} \notin \Pi_{\{1,2\}}(M_{f_{\{1,2\}}}) - M_f$ implying in particular that

$$\widehat{p} \notin \Pi_{\{1,2\}}(M_{f_{\{1,2\}}}) - M_f.$$

If $f(\overline{r}^1) = x$ then agent 2 can manipulate at $\widehat{r}^1$ towards $\overline{r}^1$, contradicting $\widehat{r}^1 \notin M_f$. Hence, $f(\overline{r}^1) \neq x$. But then agent 1 can manipulate at $\overline{r}^1$ towards p, so

$$\overline{r}^1 \in M_f.$$

Now, to end the proof of the subclaim let $p' \in \Pi_{\{1,2\}}(M_{f_{\{1,2\}}}) - M_f$ be such that p' is manipulable by $\{1,2\}$ towards q' with $|p'(N) \cup q'(N)| = 3$ and $\widehat{r}^i$ or $\overline{r}^i, i \in \{1,2\}$, is on a shortest path from p' to q'. Then $\widehat{r}^i$ $(\overline{r}^i)$ is also on a shortest path from $\widehat{p}$ to $\widetilde{p}$ $(\widehat{p}$ to $p)$ and $\widehat{p} \neq p'$ as $\widehat{p} \notin \Pi_{\{1,2\}}(M_{f_{\{1,2\}}}) - M_f$ and $p' \in \Pi_{\{1,2\}}(M_{f_{\{1,2\}}}) - M_f$. Hence, by part (2) of Lemma 48 $p' = \widetilde{p}$ $(p' = p)$. This proves the subclaim.

Now we consider three cases.

Case 1: $\{\widehat{r}^1, \widehat{r}^2\} \cap M_f = \emptyset$.

Then by the subclaim $\overline{r}^1, \overline{r}^2 \in M_f$. By the furthermore part of the subclaim $\overline{r}^1, \overline{r}^2$ can be uniquely associated to p. Then r_p^1, r_p^2 can be uniquely associated to $\widetilde{p}$.

Case 2: $\{\widehat{r}^1, \widehat{r}^2\} \cap M_f = \{\widehat{r}^i\}, i \in \{1,2\}$.

Without loss of generality $i = 1$. By the furthermore part of the claim we can uniquely associate $\overline{r}^2, r_p^1$ to p and $\widehat{r}^1, r_p^2$ to $\widetilde{p}$.

Case 3: $\{\widehat{r}^1, \widehat{r}^2\} \subset M_f$.

Now suppose, otherwise we are obviously done, that there is an $\hat{r}^i, i \in \{1,2\}$, and a $p' \in \Pi_{\{1,2\}}(M_{f_{\{1,2\}}}) - M_f$, manipulable by $\{1,2\}$ towards q' with $|p'(N) \cup q'(N)| = 3$, such that $\hat{r}^i$ is on a shortest path from p' to q'. Then, by part (2) of Lemma 48 $\{p', q'\} = \{\tilde{p}, \hat{p}\}$. If $p' = \tilde{p}$ for all such p' we are done, since we can uniquely associate $\hat{r}^1, \hat{r}^2$ to $\tilde{p}$ and r_p^1, r_p^2 to p. So, suppose that $p' = \hat{p}$ and $q' = \tilde{p}$, i.e. $\hat{p} = (\hat{t}, \hat{t}, \hat{t})$ is manipulable towards $\tilde{p} = (\tilde{t}, \tilde{t}, \tilde{t})$ by coalition $\{1,2\}$. As $f(\hat{p}) = y$ this implies that $(y, f(\tilde{p})) \in \hat{t}$ and $f(\tilde{p}) \neq y$. Since $\hat{t} = xy \ldots$ and $f(p) = x$, $p = (t, t, \tilde{t})$, it follows that $\hat{p}$ is also manipulable by $\{1,2\}$ towards p. As $\hat{p} = p' \in \Pi_{\{1,2\}}(M_{f_{\{1,2\}}}) - M_f$ this implies that $\hat{r}^1, \hat{r}^2 \in M_f$. So, we have found six manipulable profiles $\{r_p^1, r_p^2, \hat{r}^1, \hat{r}^2, \hat{r}^1, \hat{r}^2\}$ for the three distinct profiles $p, \tilde{p}, \hat{p} \in \Pi_{\{1,2\}}(M_{f_{\{1,2\}}}) - M_f$. We are done if we can show that these six manipulable profiles are uniquely associated to $p, \tilde{p}, \hat{p}$. The profiles $r_p^1, r_p^2, \hat{r}^1, \hat{r}^2$ are by construction not associated to any $p'' \in \Pi_{\{1,2\}}(M_{f_{\{1,2\}}}) - M_f, p'' \notin \{p, \tilde{p}, \hat{p}\}$. Suppose that $\hat{r}^i, i \in \{1,2\}$, is associated to $p'' \in \Pi_{\{1,2\}}(M_{f_{\{1,2\}}}) - M_f$, i.e. on a path from p'' to q''. Again then, by part (2) of Lemma 48 $\{p'', q''\} = \{p, \hat{p}\}$, and this finishes case 3.

Cases 1 to 3 show that the claim is true, and this proves the proposition by the remarks before the claim. ∎

Chapter 6

Surjectivity and Anonymity

Gibbard's (1973) and Satterthwaite's (1975) result implies that anonymous surjective so-
cial choice functions on more than two alternatives are manipulable. Placing some mild
constraints on the number of agents compared to the number of alternatives, we show what
the minimal number of manipulable profiles of such social choice functions is. Moreover,
all such social choice functions attaining the lower bound are characterized. They show a
trade off between minimizing manipulability and treating alternatives neutrally.[1]

6.1   Introduction

A well-known result of Gibbard (1973) and Satterthwaite (1975) states that any strategy-
proof surjective social choice function on more than two alternatives is dictatorial. This
implies that anonymity and surjectivity are not compatible with strategy-proofness. Here
we investigate how incompatible these are, i.e. how much manipulability we have to allow
at anonymous and surjective social choice functions.

There are a few studies on the degree of vulnerability for strategic behaviour of classical
social choice functions, such as Borda, Plurality, etc.. For example, Aleskerov and Kurbanov
(1999) contains simulation and enumeration results on 26 different social choice functions
for different indices of manipulability. Slinko (2002) counts the number of instable profiles
of classical social choice functions, which is an upper bound for the number of manipulable
profiles of these social choice functions. On the other hand, little is known about the extent
of manipulability that social choice functions satisfying certain properties must admit.
Results on this may provide lower bounds to which then all social choice functions satisfying
these properties can be compared. An investigation was pioneered by Kelly (1988), who
gave the minimal number of manipulable profiles for nondictatorial surjective social choice
functions with three alternatives and two agents and formulated several conjectures. This
line of research was continued by Fristrup and Keiding (1998), who gave the minimal
number of manipulable profiles for two agents and any number of alternatives. Maus et al.

---

[1]The results in this chapter were first formulated in Maus et al. (2004c).

(2004c, see Chapter 5) consider the three alternative case and show that for any number of agents larger than two there are six minimally manipulable nondictatorial surjective social choice functions, and these are even anonymous. However, they also consider minimally manipulable unanimous and nondictatorial social choice functions. These turn out to be nondictatorial only in a minimal sense, namely at only one profile.

Here we will exclude such social choice functions by demanding that agents are treated anonymously and we replace unanimity by surjectivity. In particular, anonymous social choice functions are nondictatorial. So, we are interested in the minimal number of manipulable profiles that an anonymous surjective social choice function has to admit. It is well known that in case of two alternatives any monotonic social choice function is strategy-proof. As monotonicity and anonymity are compatible (see e.g. May (1952)) in case of two alternatives this minimal number is zero. In case of three alternatives the minimal number is equal to the number of agents. This follows from Maus et al. (2004c, see Chapter 5). Actually, if there are more than two agents the only social choice functions reaching this minimal number are anonymous. To the best of our knowledge only so far the problem to find the minimal number of manipulable profiles of anonymous surjective social choice functions is completely solved. Although here for four or more alternatives this minimal number is determined for many cases, there are still some cases which are completely open. Due to the combinatorics used in the proofs only the following cases are solved: The four alternatives and at least five agents case and the five or more alternatives and at least two more agents than alternatives case. So, roughly speaking, the problem is not solved if there are at least as many alternatives as agents. Up till now we could neither solve the problem for any of these cases by a different proof nor could we find counterexamples. Therefore the problem is completely open for these cases.

The characterization we provide is unified for all numbers of alternatives. Interpreting the definition with two alternatives boils down to a strategy-proof social choice function for two alternatives called status-quo voting. In the three alternatives case it defines the minimally manipulable nondicatorial surjective social choice functions, as found in Maus et al. (2004c, see Chapter 5). In general the characterized social choice functions show how one can (but probably would not want to) minimize manipulability while treating agents anonymously. They minimize also surjectivity, meaning that except for two alternatives, say b and c, for all other alternatives there is precisely one profile at which they are chosen. At the remaining profiles the choice between b and c is on a unanimity basis: c is chosen unless b is unanimously preferred to b. So, alternatives are treated in a nonneutral way, which will usually be just as undesirable as treating agents unequal. Nevertheless the result shows what is possible if we want to minimize manipulability.

It should be noted that the way in which we measure manipulability, namely by counting manipulable profiles, is not the only possible one. In fact there are a lot of variations of this, where for example profiles are counted by the number of agents that can manipulate them, or the severity of the manipulation measured by the distance between what the agent

achieves with and without manipulation is taken into account. Aleskerov and Kurbanov (1999) contain an overview of such variations and numerical results on the manipulability of 26 social choice functions according to these measurements.

Within the restrictions pointed out we show as a main step of our proof of the main result that strategy-proof and anonymous surjective social choice between more than two alternatives is not possible on subsets of the whole domain satisfying a certain diversity condition, namely on subsets where agents hold at least $k \in \mathbb{N}$ different preferences. We point this out here since this is an impossibility result on a restricted domain that is of interest on its own. It says, that the impossibility result of Gibbard-Satterthwaite cannot be resolved by demanding diversity of preferences. The reader who is familiar with a proof of Arrow's theorem (Arrow, 1963) and a proof Gibbard-Satterthwaite's theorem building on Arrow's theorem will recognize a lot of similarities. The basic idea is to show that the steps taken there can be adopted to go through also on the restricted domains considered here. To ensure that these steps can be adopted we need the restrictions that then cause the restrictions on the main result.

In the next section the relevant definitions are given, and the relation to Gibbard (1973), Satterthwaite (1975) and Maus et al. (2004c, see Chapter 5) is made more precise. Sections 6.4 and 6.5 link strategy-proofness to monotonicity, Pareto optimality and decisiveness on sets of profiles where agents hold at least k different preferences. Then, Section 6.6 combines this to show that there are no strategy-proof and anonymous surjective social choice functions selecting among more than two alternatives on these restricted domains. After that, Section 6.7 characterizes the minimally manipulable anonymous surjective social choice functions selecting from more than three alternatives.

6.2 Preliminaries

We denote the cardinality of a set S by $|S|$.

Let A be a finite set of alternatives, $m := |A| \geq 3$, and N a finite set of agents, $n := |N| \geq 2$.

Let $t \subset A \times A$. We call t *complete* if for all $x, y \in A$ $(x, y) \in t$ or $(y, x) \in t$. Note that completeness of t implies $(x, x) \in t$ for all $x \in A$. We call t *transitive* if for all $x, y, z \in A$ $(x, y) \in t$ and $(y, z) \in t$ implies $(x, z) \in A$. We call t *antisymmetric* if for all $x, y \in A$ $(x, y) \in t$ and $(y, x) \in t$ implies that $x = y$.

A *preference* $t \subset A \times A$ is a *linear ordering* (complete, transitive, antisymmetric) on A. Let P denote the set of all preferences. Suppose that $A = \{x_1, x_2, \ldots x_m\}$. By completeness, transitivity and antisymmetry we can write conveniently

$$t = x_1 x_2 \ldots x_m$$

for the preference t such that $(x_i, x_j) \in t$ if and only if $i \geq j$, $i, j \in \{1, 2, \ldots, m\}$, and

$$t = \ldots x \ldots y \ldots$$

if we want to express only that x is strictly preferred to y.

A *profile* p is a map from N to P. Let P^N denote the set of all these maps. Thus, a profile assigns to every agent i a preference $p(i)$ over the alternatives. For a nonempty subset S of N we denote by $p|_S$ the *restriction* of the map p to the domain S. We denote a profile such that all agents have the same preference $t \in P$ by t^N.

A *social choice function* is a function $f : P^N \to A$. Hence, a social choice function selects a unique alternative $f(p)$ at every profile p.

A social choice function is called *surjective* if any alternative in A is chosen at least once, i.e. if $f(P^N) = A$. In the literature this is also known as citizen-sovereignty. Throughout this chapter we assume that any social choice function is surjective, and do not mention this anymore explicitly.

For a permutation σ of N and a profile $p \in P^N$ let $p \circ \sigma$ be the profile given by $(p \circ \sigma)(i) := p(\sigma(i))$ for all $i \in N$. A social choice function is called *anonymous* if $f(p) = f(p \circ \sigma)$ for all permutations σ of N. Thus, anonymous social choice functions are symmetric in the arguments. In a sense they treat agents equally.

In contrast to anonymity, the following dictatorial social choice functions $dict_d$ respect only the preference of one single agent $d \in N$, the *dictator*. For any profile p $dict_d$ is defined by

$$dict_d(p) := x$$

where x is such that $p(d) = x \dots$. So, $dict_d(p)$ is the most preferred alternative of agent d in $p(d)$. A social choice function f is called *nondictatorial* if there is no agent d such that $f = dict_d$.

6.3 Manipulation of social choice functions

We are interested in strategic behaviour of individuals when facing cooperative decision-making as captured by social choice functions. This is formalized by the following definitions. Let $f : P^N \to A$ be a social choice function. Let $p \in P^N$ be a profile. Then each profile q such that, for some $i \in N$, $q|_{N-\{i\}} = p|_{N-\{i\}}$ and $q(i) \neq p(i)$, is called an *i-deviation* from p. Letting $t := q(i)$ we use the notation $q = (p_{-i}, t)$. If it is not important which agent deviates from p to q we call q a *deviation* from p. A profile p is called *manipulable* (under f) if there is an agent that is better off by being dishonest about his preference, i.e. if there is an $i \in N$, and an i-deviation q such that

$$(f(p), f(q)) \notin p(i).$$

In this case we say that p is *manipulable towards* q (under f). Let $B \subseteq P^N$ and

$$M_f(B) \ := \ \{p \in B \mid p \text{ is manipulable towards}$$
$$\text{some } q \in B \text{ under } f\}.$$

A social choice function is called *strategy-proof* on B if $M_f(B) = \emptyset$, otherwise it is said to be *manipulable on B*. When $B = P^N$ we omit B and write $M_f := M_f(P^N)$.

The prominence of the dictatorial rules arises from the following impossibility result due to Gibbard (1973) and Satterthwaite (1975).

Theorem 54 *Let A be a finite set of alternatives, $|A| \geq 3$. Let $f : P^N \rightarrow A$ be a nondictatorial social choice function. Then*

$$|M_f| \geq 1.$$

However, this theorem makes no statement about the number of manipulable profiles such social choice functions admit. This question has been solved by Kelly (1988) for two agents and three alternatives, and by Fristrup and Keiding (1998) for two agents, and any number of alternatives larger than three. In Maus et al. (2004c, see Chapter 5) it is shown that for $m = 3$ and any number of agents larger than two, the following social choice functions are exactly the minimally manipulable ones among the nondictatorial social choice functions.

Definition 55 *Choose $b, c \in A, b \neq c$, and let $A = \{a_1, \ldots a_{m-2}, b, c\}$. Choose $m - 2$ different preferences $t_j \in P, j \in \{1, 2, \ldots, m - 2\}$, such that $t_j = \ldots a_j \ldots b \ldots c \ldots$ for all $j \in \{1, 2, \ldots, m - 2\}$. Let $m : P^N \rightarrow A$ be the social choice function given by*

$$m(p) := \begin{cases} a_j & \text{if } p = t_j^N \text{ for some } j \in \{1, 2, \ldots, m - 2\} \\ b & \text{if } (b, c) \in p(i) \text{ for all } i \in N \text{ and } p \notin \{t_1^N, \ldots t_{m-2}^N\} \\ c & \text{if there is an } i \in N \text{ such that } (c, b) \in p(i). \end{cases}$$

Note that the social choice functions given by Definition 55 are not only nondictatorial but even anonymous. So, they are the minimally manipulable anonymous social choice functions for three alternatives. Hence, we do not have to treat the three alternative case here. We will show that also for $m \geq 4$ the social choice functions given by Definition 55 are the minimally manipulable anonymous social choice functions. One might hope that as in the case $m = 3$ the minimally manipulable nondictatorial social choice functions coincide with the anonymous ones. We will see that this is not so.

6.4 Monotonicity and Pareto optimality on B^k

Let $k \in \mathbb{N}$ and let

$$B^k := \{p \in P^N \mid |p^{-1}(N)| \geq k\}$$

be the set of all profiles that contain at least k different preferences. We want to show an impossibility result for strategy-proof and anonymous social choice functions on B^k for certain k. This section contains two ingredients for that impossibility result, a monotonicity lemma and a Pareto optimality lemma.

Let $p, q \in P^N, a \in \tilde{A} \subseteq A$. We say that *a weakly improved* from p to q with respect to $\tilde{A}$, if for all alternatives $x \in \tilde{A}$, we have for all agents $i \in N$

$$(a, x) \in p(i) \text{ implies } (a, x) \in q(i).$$

Let $f : P^N \to A$ be a social choice function and let $B \subseteq P^N$. We say that f is *monotone* on B, if for all $p, q \in B$ such that $f(p)$ weakly improved from p to q with respect to $f(B)$, we have

$$f(p) = f(q).$$

This monotonicity condition corresponds to the one used in Dasgupta et al. (1979) for social choice rules, and is also known as Maskin-monotonicity. The following lemma links strategy-proofness to monotonicity.

Lemma 56 *Let $f : P^N \to A$ be anonymous and strategy-proof on $B^k, k \in \mathbf{N}$. Then f is monotone on B^k.*

Proof. Let $p, q \in B^k$ be such that $f(p)$ weakly improves from p to q with respect to $f(B^k)$. We have to show that $f(p) = f(q)$.

Claim 1: If there are $i, j \in N, i \neq j$, such that $p(i) = q(j)$, then $f(p)$ also weakly improves from p to $\bar{q} := q \circ (i, j)$ with respect to $f(B^k)$, where $(i \ j)$ is the permutation mapping i to j and j to i, and leaving all other elements of N fixed.

For all $l \in N - \{i, j\}$ we have that

$$(f(p), x) \in p(l) \text{ implies } (f(p), x) \in \bar{q}(l) \text{ for all } x \in f(B^k), \tag{6.1}$$

as $\bar{q}(l) = q(l)$ and $f(p)$ weakly improves from p to q with respect to $f(B^k)$. For $l = i$ we have $p(l) = p(i) = q(j) = \bar{q}(i) = \bar{q}(l)$, so that statement 6.1 also holds for $l = i$. For $l = j$ and all $x \in f(B^k)$ we have $(f(p), x) \in p(l)$ implies $(f(p), x) \in q(l) = q(j) = p(i)$, as $f(p)$ weakly improves from p to q with respect to $f(B^k)$. Then for the same reason $(f(p), x) \in p(i)$ implies that $(f(p), x) \in q(i) = \bar{q}(j) = \bar{q}(l)$. Altogether, statement 6.1 holds for all $l \in N$, so $f(p)$ weakly improves from p to $\bar{q}$ with respect to $f(B^k)$. This proves claim 1.

Note that $p(i) = \bar{q}(i)$ for the $\bar{q}$ constructed in claim 1 and that $\bar{q} \in B^k$. By repeatedly appying claim 1 we obtain a $\bar{q} \in B^k$ such that $p(i) \neq \bar{q}(j)$ for all $i, j \in S := \{i \in N | p(i) \neq \bar{q}(i)\}$, i.e. $p(S) \cap q(S) = \emptyset$. By anonymity $f(q) = f(\bar{q})$, so it suffices to show that $f(\bar{q}) = f(p)$. Furthermore, by anonymity we may assume that $S = \{1, 2, \ldots, s\}, s = |S|$, and we may number the agents in S such that for some largest possible $v \in \{1, 2, \ldots, s\}$ we have that $q(j) \notin q((N - S) \cup \{1, \ldots, j - 1\})$ for all $j \in \{1, \ldots, v - 1\}$.

Let the profiles $r^l \in P^N, l \in \{0, \ldots, n\}$, be given by

$$r^l(i) := \begin{cases} \bar{q}(i) & i \leq l \\ p(i) & i > l \end{cases}$$

for all $i \in N$.

Claim 2: $r^l \in B^k$ for all $l \in \{0, \dots, s\}$.

We prove the claim by induction on j.

Induction Basis: $j = 0$. Since $r^0 = p \in B^k$ this is obvious.

Induction step: Suppose that $|r^{j-1}(N)| \geq k$ for some $j \in \{1, \dots, s\}$. We have to show that $|r^j(N)| \geq k$. If $j < v$, then $\bar{q}(j) \notin \bar{q}((N - S) \cup \{1, \dots, j-1\})$. Furthermore, as $p(S) \cap \bar{q}(S) = \emptyset$ and $j \in S$ we have $\bar{q}(j) \notin p(\{j+1, \dots, s\})$. Since

$$
\begin{aligned}
r^j(N) &= \bar{q}(\{1, \dots, j-1\}) \cup \{\bar{q}(j)\} \cup p(\{j+1, \dots, s\}) \cup p(N - S) \\
&= \bar{q}(\{1, \dots, j-1\}) \cup \{\bar{q}(j)\} \cup p(\{j+1, \dots, s\}) \cup \bar{q}(N - S)
\end{aligned}
$$

it follows that $\bar{q}(j) \notin r^j(N - \{j\})$. As $r^{j-1}(N - \{j\}) = r^j(N - \{j\})$, $\bar{q}(j) \notin r^j(N - \{j\})$ and $|r^{j-1}(N)| \geq k$, hence $|r^{j-1}(N - \{j\})| \geq k - 1$, it follows that $|r^j(N)| \geq k$. To end the proof of the induction step suppose $j \geq v$. By the maximality of v we have $\bar{q}(i) \in \bar{q}((N - S) \cup \{1, \dots, v-1\})$ for all $i \in \{v, \dots, s\}$. So, $\bar{q}(N) = \bar{q}((N - S) \cup \{1, \dots, v-1\})$. But then

$$
\begin{aligned}
\bar{q}(N) &= \bar{q}(S) \cup \bar{q}(N - S) \\
&= \bar{q}(\{1, \dots, v-1\}) \cup \bar{q}(\{v, \dots, s\}) \cup \bar{q}(N - S) \\
&= \bar{q}(\{1, \dots, v-1\}) \cup \bar{q}(N - S) \\
&= r^j(\{1, \dots, v-1\}) \cup r^j(N - S) \\
&\subseteq r^j(N),
\end{aligned}
$$

and hence $\bar{q} \in B^k$ implies that $r^j \in B^k$. This finishes the proof of claim 2.

We can now show that $f(\bar{q}) = f(p)$. First of all $r^0 = p$ and $r^n = \bar{q}$. By claim 2 $r^l \in B^k$ for all $l \in \{0, \dots, n\}$. As f is strategy-proof on B^k we have $(f(r^1), f(r^0)) \in r^1(1) = \bar{q}(1)$ and $(f(r^0), f(r^1)) \in p(1)$. As $f(r^0) = f(p)$ and $f(p)$ weakly improved from p to $\bar{q}$ with respect to $f(B^k)$, $(f(r^0), f(r^1)) \in p(1)$ implies $(f(r^0), f(r^1)) \in \bar{q}(1)$. So,

$$
(f(r^1), f(r^0)) \in r^1(1) = \bar{q}(1)
$$

and

$$
(f(r^0), f(r^1)) \in \bar{q}(1).
$$

This implies that $f(p) = f(r^0) = f(r^1)$, since preferences are antisymmetric. Repeating this argument yields

$$
f(p) = f(r^1) = f(r^2) = \dots = f(r^n) = f(\bar{q}),
$$

which finishes the proof. ∎

As a corollary of Lemma 56 we can state the following.

Corollary 57 *Let $f : P^N \to A$ be anonymous and strategy-proof on $B^k, k \in \mathbb{N}$. Let $a \in$ $f(B^k)$ and $p \in B^k$ be such that $(x, a) \notin p(i)$ for all $i \in N$ and all $x \in f(B^k) - \{a\}$. Then $f(p) = a$.*

Let $x, y \in A, x \neq y$. We say that x *Pareto dominates* y at the profile p, if $(x, y) \in p(i)$ for all $i \in N$. A social choice function is called *Pareto optimal* on $B \subseteq P^N$, if it does not choose alternatives Pareto dominated by an element of $f(B)$, i.e. there is no $x \in f(B) - \{f(p)\}$ such that x Pareto dominates $f(p)$ at p.

The second lemma shows that if f is anonymous and strategy-proof on B^k, then f is Pareto optimal on B^k. We have to make restrictions on k in order to make sure that there are sufficiently many different preferences having the same alternative at the first and second places.

Lemma 58 *For $m = 4$ let $k \leq m + 1$ and for $m \geq 5$ let $k \leq m + 2$. Let $f : P^N \to A$ be strategy-proof and anonymous on B^k. Then f is Pareto optimal on B^k.*

Proof. Let $p \in B^k, x \in f(B^k)$ and $y \in A$ be such that x Pareto dominates y at p. Suppose that $f(p) = y$. It is sufficient to deduce a contradiction. By Lemma 56 f is monotone on B^k. As $p \in B^k$ and f is anonymous we can without loss of generality suppose that $|p(\{1, \ldots, k\})| = k$. Let $l := \max\{k - (m - 2)!, 0\}$, then by monotonicity we can assume that for $l < i \leq k$ we have $p(i) = xy \ldots$. In view of Corollary 57, $l \geq 1$. For $i \in \{1, \ldots, l\}$ let $Z_i := \{a \in A \mid (a, y) \in p(i)\}$ be the upper contour of y at $p(i)$. There are precisely $(m - 2)!$ preferences where x is best and y is second best and $l \geq 1$, hence $k > (m - 2)!$ and $\{x, y\} \subsetneq Z_i$ for all $i \in \{1, \ldots, l\}$. Consider $r(i)$ obtained from $p(i)$ by shifting alternative x to the top leaving all other alternatives unchanged:

$$r(i) = (\{x\} \times A) \cup \left(p(i) \cap (A - \{x\})^2\right).$$

For $l < i \leq k$ we have $r(i) = p(i)$. If we show that $r \in B^k$ then by monotonicity $f(r) = y$, as $y = f(p)$ improves from p to r.

Case 1: $|\{Z_1, \ldots, Z_l\}| = l$.

As $\{x, y\} \subsetneq Z_i$ and $|\{Z_1, \ldots, Z_l\}| = l$, $|r(\{1, \ldots, k\})| = k$. Hence, $r \in B^k$. By Corollary 57 $f(r) = x$, contradicting $f(r) = y$.

Case 2: $|\{Z_1, \ldots, Z_l\}| < l$.

So, $l \geq 2$. But $l = k - (m - 2)! \leq (m + 2) - (m - 2)! \leq 1$ for $m \geq 5$. So, $m = 4$ and $l \in \{2, 3\}$. Without loss of generality $Z_1 = Z_2$ and let $A = \{x, y, a, b\}$. As $\{x, y\} \subsetneq Z_i$ for all $i \in \{1, \ldots, l\}$, there are alternatives $z_i \in Z_i - \{x, y\}$ and, if $l = 3$, $z_3 \in Z_3 - \{x, y\}$. If $l = 3$ and $Z_3 = Z_1 = Z_2$ we must have $|Z_1| = 4$ as $|Z_1| = 3$ implies that without loss of generality $p(\{1, 2, 3\}) \subseteq \{axyb, xayb\}$, contradicting $|p(\{1, \ldots, k\})| = k$. So, if $l = 3$, we can choose a $z_3 \in Z_1 - \{x, y, z_1\}$. Without loss of generality suppose that $z_1 = a$ and, if

$l = 3$, $z_3 = b$. Consider the profile $\bar{r}$ defined by

$$\bar{r}(i) := \begin{cases} axyb & \text{if } i = 1, \\ xayb & \text{if } i = 2, \\ xbya & \text{if } i = l = 3, \\ r(i) & \text{if } i > l. \end{cases}$$

Then $\bar{r} \in B^k$ and y improves from p to $\bar{r}$. Hence, $f(\bar{r}) = y$. Let $\hat{r} := (r_{-1}, xaby)$. By Lemma 57, $f(\hat{r}) = x$. So, $\hat{r}$ is a 1-deviation from $\bar{r}$ such that $(f(\bar{r}), f(\hat{r})) \notin \bar{r}(1)$. But then $\bar{r} \in B^k$ is manipulable and this contradiction ends the proof. $\blacksquare$

6.5   Decisiveness on $B^k_{\{a,b\}}$

Let $X \subseteq A$ and B^k_X be the set of all profiles in B^k where all alternatives in X are strictly preferred to alternatives in $A - X$,

$$B^k_X := \{p \in B^k \mid \text{for all } i \in N \text{ and } x \in X, y \in A - X \text{ we have } (x,y) \in p(i)\}.$$

Let $p \in B^k_{\{a,b\}}$, $a \neq b$. We say that $S \subseteq N$ decides over (a,b) at p, if $f(p) = a$ and $S = \{i \in N \mid (a,b) \in p(i)\}$. We say that S decides over (a,b), if S decides over (a,b) at all profiles $p \in B^k_{\{a,b\}}$ that satisfy $S = \{i \in N \mid (a,b) \in p(i)\}$. We denote the set of all S that are decisive over (a,b) by $W^k(a,b)$. The following is an immediate consequence of the definitions of weak improvement and monotonicity.

Remark 59 *For social choice functions f which are monotone on B^k, S decides over (a,b) at a profile $p \in B^k_{\{a,b\}}$ if and only if $S \in W^k(a,b)$.*

The next lemma shows some decisiveness properties of f on B^k over triples of alternatives, when f is strategy-proof on B^k.

Lemma 60 *Let $k \leq m$ if $m = 4$ and $k \leq m + 2$ if $m \geq 5$. Let f be anonymous and strategy-proof on B^k, $f(B^k) \supseteq \{a,b,c\}$ for different a, b and c. Let $S \in W^k(a,b)$. Then*

1. $S \in W^k(c,b)$ and $S \in W^k(a,c)$,

2. $S \in W^k(b,a)$ and

3. $|S| > \frac{1}{2}n$.

Proof. (1): By Lemma 56 f is monotone on B^k. By the restrictions on k $B^k_{\{a,b\}}$ is nonempty. Let $p \in B^k_{\{a,b\}}$ be such that S is decisive over (a,b) at p. For a permutation τ on A and a preference $t \in P$ let

$$t^\tau = \{(\tau(x), \tau(y)) \mid (x,y) \in t\}.$$

For $x, y \in A$ let $(x\ y)$ denote the permutation that maps x on y, y on x and all other elements of A on themselves.

Let p^1 be such that $p^1|_S = p|_S$ and $p^1(i) = (p(i))^{(a\ c)}$ for all $i \in N - S$. Then $p^1 \in B^k$. By monotonicity $f(p^1) \neq b$, as b weakly improves from p^1 to p and $f(p) = a \neq b$. As b Pareto dominates all $x \in A - \{a, b\}$ at p^1 we also have by Lemma 58 that $f(p^1) \notin A - \{a, b\}$, so $f(p^1) = a$.

Let p^2 be such that $p^2|_{N-S} = p^1|_{N-S}$ and $p^2(i) = (p^1(i))^{(b\ c)}$ for all $i \in S$. Then $p^2 \in B^k$. By monotonicity $f(p^2) = a$, as a weakly improves from p^1 to p^2 and $f(p^1) = a$.

Let p^3 be such that $p^3|_{N-S} = p^2|_{N-S}$ and $p^3(i) = (p^2(i))^{(a\ c)}$ for all $i \in S$. Then $p^3 \in B^k$. As $c \in f(B^k)$ Pareto dominates all alternatives $x \in A - \{b, c\}$ it follows by Lemma 58 that $f(p^3) \in \{b, c\}$. By monotonicity $f(p^3) \neq b$, as b weakly improves from p^3 to p^2 and $f(p^2) = a \neq b$. So, $f(p^3) = c$.

Finally, let p^4 be such that $p^4|_{N-S} = p^3|_{N-S}$ and $p^4(i) = (p^3(i))^{(a\ b)}$ for all $i \in S$. Then $p^4 \in B^k_{\{b,c\}}$ and by monotonicity $f(p^4) = c$. So, S decides over (c, b) at p^4 and by Remark 59

$$S \in W^k(c, b).$$

In order to see that S decides over (a, c) let p^5 be such that $p^5|_S = p^2|_S$ and $p^5(i) = (p^2(i))^{(b\ c)}$ for all $i \in N - S$. Then $p^5 \in B^k$ and as $f(p^2) = a$ we have by monotonicity that $f(p^5) = a$.

Let p^6 be such that $p^6|_S = p^5|_S$ and $p^6(i) = (p^5(i))^{(a\ b)}$ for all $i \in N - S$. Then $p^6 \in B^k_{\{a,c\}}$ and by monotonicity $f(p^6) = a$. So, S decides over (a, c) at p^6 and by Remark 59

$$S \in W^k(a, c).$$

(2): By (1) we have $S \in W^k(a, c)$. Thus, applying (1) to $S \in W^k(a, c)$ yields $S \in W^k(b, c)$ and then applying the second part of (1) to $S \in W^k(b, c)$ yields $S \in W^k(b, a)$.

(3): To the contrary suppose that $|S| \leq \frac{1}{2}n$. By anonymity we have $T \in W^k(a, b)$ for all T such that $|T| = |S|$. By (2) also $T \in W^k(b, a)$ for all such T. By monotonicity then $T \in W^k(b, a)$ for all T such that $|T| \geq |S|$. So, $N - S \in W^k(b, a)$ and this contradicts $S \in W^k(a, b)$. ∎

In the case $m = 4$ and $k = m + 1 = 5$, $B^k_{\{a,b\}}$ is empty, but an almost similar result can be achieved using the following terminology. Let $A = \{a, b, c, d\}$. Let $B^{k,d}_{\{a,b\}}$ be the set of all profiles p in B^k where all except one agent $i \in N$ prefer all alternatives in $\{a, b\}$ to all alternatives in $\{c, d\}$, and $p(i) \in \{adbc, bdac\}$. Let $p \in B^{k,d}_{\{a,b\}}$, $a \neq b$. We say that $S \subseteq N$ d-decides over (a, b) at p, if $f(p) = a$ and $S = \{i \in N \mid (a, b) \in p(i)\}$, $adbc \notin p(S)$. We say that S d-decides over (a, b), if S d-decides over (a, b) at all profiles $p \in B^{k,d}_{\{a,b\}}$ that satisfy $S = \{i \in N \mid (a, b) \in p(i)\}$, $adbc \notin p(S)$. We denote the set of all S that are d-decisive over (a, b) by $W^k_d(a, b)$.

As before we make the following observation.

Remark 61 *For social choice functions f, which are monotone on B^k, S d-decides over (a, b) at a profile $p \in B^{k,d}_{\{a,b\}}$ if and only if $S \in W^k_d(a, b)$.*

Now we can show a lemma like Lemma 60 for d-decisiveness.

Lemma 62 *Let $A = \{a, b, c, d\}$ and $k = 5$. Let f be anonymous and strategy-proof on B^k, $f(B^k) \supseteq \{a, b, c\}$ for pairwise different a, b and c. Let $S \in W^k_d(a, b)$. Then*

1. $S \in W^k_d(c, b)$ and $S \in W^k_d(a, c)$

2. $S \in W^k_d(b, a)$

3. $|S| \geq \frac{1}{2}n$.

Proof. By Lemma 56 f is monotone on B^k. Let $S \in W^k_d(a, b)$. Then $|p(S)| = 2$ by definition, and thus $2 \leq |S| \leq n - (k - |p(S)|) = n - 3$. Without loss of generality $S = \{1, \ldots, |S|\}$. Consider the following profiles.

	p	p^1	p^2	p^3	p^4	p^5	p^6
$\{1\}$	abcd	abcd	acbd	cabd	cbad	acbd	acbd
$S - \{1\}$	abdc	abdc	acdb	cadb	cbda	acdb	acdb
$N - \{n-1, n\}$	bacd	bcad	bcad	bcad	bcad	cbad	cabd
$n - 1$	badc	bcda	bcda	bcda	bcda	cbda	cadb
n	bdac	bdca	bdca	bdca	bdca	cdba	cdab

As in the proof of Lemma 60 by these profiles it follows that $S \in W^k_d(c, b)$ and $S \in W^k_d(a, c)$. Then (2) follows by applying (1) twice as in the proof of Lemma 60. Furthermore, if $T \subseteq N$ such that $|T| \geq S$ and $|T| \leq n - 3$, then by anonymity and monotonicity $T \in W^k_d(b, a)$. Suppose $S \in W^k_d(a, b)$ and $|S| < \frac{1}{2}n$. Then $|N - (S \cup \{n\})| = |N| - |S| - 1 > \frac{1}{2}n - 1 \geq |S|$. So, $N - (S \cup \{n\}) \in W^k_d(b, a)$. Consider the profile $q = (p_{-n}, adbc)$. By strategy-proofness $f(q) = a$. But $\{i \in N \mid (b, a) \in q(i)\} = N - (S \cup \{n\}) \in W^k_d(b, a)$, so $f(q) = b$, a contradiction. ∎

6.6 Gibbard-Satterthwaite on B^k

We combine the results of the previous sections to show an impossibility result for strategy-proof anonymous social choice functions among more than two alternatives on the restricted domain B^k, i.e. when agents preferences have to satisfy a minimal diversity.

Theorem 63 *If $m = 4$ let $k \leq m + 1 \leq n$ and if $m \geq 5$ let $k \leq m + 2 \leq n$. Let f be strategy-proof on B^k and anonymous. Then $|f(B^k)| \leq 2$.*

Proof. To the contrary suppose that there are pairwise different $a, b, c \in f(B^k)$. Note that by the constraints $n \geq 5$. By Lemma 56 f is monotone on B^k. Let $\{S, T, U\}$ be a partition of N, such that for $X, Y \in \{S, T, U\}$ we have $0 < |X| \leq |Y| + 1$. Consider profiles p such that

$$p(S) \subseteq \{t \in P \mid t = ab\ldots\},$$
$$p(T) \subseteq \{t \in P \mid t = bc\ldots\}, \text{ and}$$
$$p(U) \subseteq \{t \in P \mid t = ca\ldots\}.$$

There are $(m-2)!$ different choices for any $p(i)$, hence we may assume that $p \in B^k$. Without loss of generality $f(p) = a$. For some $\tilde{S}$ such that $|\tilde{S}| < \frac{1}{2}n$ we will prove that $\tilde{S} \in W^k(a, c)$ if $(m, k) \neq (4, 5)$ and $\tilde{S} \in W_d^k(a, c)$ if $(m, k) = (4, 5)$. As $|\tilde{S}| < \frac{1}{2}n$, we have a contradiction with Lemma 60, or Lemma 62 respectively, and are done.

Case 1: $(m, k) \neq (4, 5)$

Let $\{\tilde{S}, \tilde{T}, \tilde{U}\}$ be a partition of N such that $S \subseteq \tilde{S}$, $T \supseteq \tilde{T}$, $U \supseteq \tilde{U}$ and for $X, Y \in \{\tilde{S}, \tilde{T}, \tilde{U}\}$ we have $0 < |X| \leq \min\{|Y| + 1, |\tilde{S}|\}$. Consider q such that

$$q(\tilde{S}) \subseteq \{t \in P \mid t = ac\ldots\},$$
$$q(\tilde{T}) \subseteq \{t \in P \mid t = ca\ldots\} \text{and}$$
$$q(\tilde{U}) \subseteq \{t \in P \mid t = ca\ldots\}.$$

A preference $q(i)$ in $\tilde{S}$ or $\tilde{T} \cup \tilde{U}$ can be extended in $(m-2)!$ ways. We prove that there are such $q \in B_{\{a,c\}}^k$. So, we have to show that

$$k \leq \min\{|\tilde{S}|, (m-2)!\} + \min\{|\tilde{T}| + |\tilde{U}|, (m-2)!\} =: \alpha.$$

There are four cases. If $\alpha = |\tilde{S}| + |\tilde{T}| + |\tilde{U}| = |N|$ then $\alpha = n \geq k$. If $\alpha = |\tilde{S}| + (m-2)!$, then, as $|\tilde{S}| \geq 2$,

$$\alpha \geq 2 + (m-2)! \geq m + 2 \geq k$$

if $m \geq 5$, and

$$\alpha \geq 2 + (m-2)! = m \geq k$$

if $m = 4$, since $(m, k) \neq (4, 5)$. If $\alpha = 2(m-2)!$, then $\alpha \geq 2 + (m-2)!$ and we can argue as in the previous case. As $n \geq m + 1$ and $m \geq 4$, $n \geq 5$. So, $|\tilde{S}| < |\tilde{T}| + |\tilde{U}|$. Therefore the case $\alpha = (m-2)! + |\tilde{T}| + |\tilde{U}|$ cannot occur. So, we may assume that $q \in B_{\{a,c\}}^k$.

From p to q a weakly improves, so by monotonicity $f(q) = a$, and therefore $\tilde{S} \in W^k(a, c)$. By the way $\tilde{S}$ was chosen we have $|\tilde{S}| < \frac{1}{2}n$, contradicting part 3 of Lemma 60.

Case 2: $(m, k) = (4, 5)$

Subcase 2^A : $|S| = 1$

Then $n = 5$, and by anonymity we may assume that

$$p = (ac\ldots, bcad, bcda, cabd, cadb).$$

Let
$$q = (acdb, acbd, cdab, cabd, cadb).$$

Then a has improved from p to q, so $f(q) = a$ by monotonicity. But then $\{i \in N \mid (a,c) \in q(i)\} = S = \{1,2\} \in W_d^k(a,c)$, contradicting part 3 of Lemma 62.

Subcase 2^B : $|S| \geq 2$

Let p^1 be such that $p^1|_{N-S} = p|_{N-S}$ and $p^1(i) = (p(i))^{(b\,c)}$ for all $i \in S$. Then $p^1 \in B^k$. By monotonicity $f(p^1) = a$, as a weakly improves from p to p^1. Without loss of generality $1 \in T$. Let p^2 be equal to p^1 if there is an agent $\hat{i} \in T$ with $p^1(i) = bcda$. So, $f(p^2) = a$. Otherwise $p^1(T) = \{bcad\}$ and we let $p^2 := (p_{-\hat{i}}, bcda)$ for some $\hat{i} \in T$. Obviously $p^2 \in B^5$. Now c Pareto dominates d at p^2, so $d \neq f(p^2)$ by Lemma 58. As f is strategy-proof on B^5 and $f(p^2) \notin \{b,c\}$ we obtain $f(p^2) = a$ in the case $p^1(T) = \{bcad\}$ as well. Now consider p^3 such that $p^3|_{N-T} = p^2|_{N-T}$, $p^3(1) = cdab$ and $p^3(T - \{1\}) = \{cabd\}$. Then a weakly improves from p^2 to p^3, and $p^3 \in B^k$, so by monotonicity $f(p^3) = a$. But then $\{i \in N \mid (a,c) \in p^3(i)\} = S \in W_d^k(a,c)$, contradicting part 3 of Lemma 62. ∎

The theorem says that within the restrictions the impossibility result of Gibbard Satterthwaite holds already on the smaller domain of profiles that satisfy a certain diversity. This may not surprise the reader because practically situations with a lot of diversity can be expected to be suspect to manipulation. At least it confirms this expectation. We did not think much about the situation when k exceeds the bounds given. This is beyond the scope of this chapter since the result is only a by-product on the way to the main result contained in the next section.

6.7 Minimal manipulability of anonymous social choice functions

Before proving the main result we show some helpful inequalities.

Lemma 64 Let $k, l, m, n \in \mathbb{N}$. Then

1. $k!l! \leq (k+l-1)!$

2. If $2 \leq k \leq m+1 \leq n$ and $m \geq 4$, then

$$\left(\frac{m!}{3} - k\right) \frac{(n-1)!}{(n-k)!} > \left(\frac{m!}{3} - 1\right)(m-2).$$

Proof. (1):

$$k!l! \leq (k+l-1)! \iff k+l \leq \binom{k+l}{k} \iff k, l \geq 1$$

(2): Note that

Proof.

$$\left(\tfrac{m!}{3} - k\right) \tfrac{(n-1)!}{(n-k)!} > \left(\tfrac{m!}{3} - 1\right)(m-2)$$

$$\iff \left(\tfrac{m!}{3} - k\right)\left(\tfrac{(n-1)!}{(n-k)!(m-2)}\right) - \tfrac{m!}{3} > -1$$

$$\iff \tfrac{m!}{3}\left(\left(\tfrac{(n-1)!}{(n-k)!(m-2)}\right) - 1\right) > k\left(\tfrac{(n-1)!}{(n-k)!(m-2)}\right) - 1$$

Let $\alpha := \tfrac{(n-1)!}{(n-k)!(m-2)}$. Then we have to show that $\tfrac{m!}{3}(\alpha - 1) > k\alpha - 1$.

Now, for $m \geq 5$ we have $m! \geq 3(m+1)^2$, so it is sufficient to prove that $(m+1)^2(\alpha-1) \geq k\alpha$. In view of $k \leq m+1$ it is then enough to show that $(m+1)\alpha - 1 \geq \alpha \iff m\alpha > 1$ which is satisfied, as clearly $\alpha > \tfrac{1}{m-2}$. If $m = 4$ then we have to show that $8(\alpha - 1) > k\alpha - 1 \iff (8 - k)\alpha > 7$.

If $k = 2$ then this becomes $6\tfrac{(n-1)}{2} > 7$, which is satisfied as $n \geq 5$. If $2 < k \leq 5$ then also $(8 - k)\alpha > 3\tfrac{(n-1)(n-2)}{2} > 7$, as $n \geq 5$. This proves (2). ∎

Now we can prove the main result. It characterizes the minimally manipulable anonymous social choice functions for more than three alternatives.

Theorem 65 *Let $f : P^N \to A, m \geq 4$, be an anonymous social choice function. Suppose that $n \geq m + 1$ if $m = 4$ and that $n \geq m + 2$ if $m \geq 5$. Then*

$$|M_f| \geq n\left(\frac{m!}{3} - 1\right)(m - 2).$$

Furthermore, equality holds if and only if f can be described as in Definition 55.

Proof. Let $P = \{t_1, \ldots, t_{m!}\}$ and $0! = 1$. Let $p \in P^N$ and let $c(p, i) := |p^{-1}(t_i)|$, $i \in \{1, \ldots, m!\}$. Then there are

$$\frac{n!}{\prod_{i=1}^{m!} c(p, i)!}$$

profiles $q \in P^N$ that satisfy $c(q, i) = c(p, i)$ for all $i \in \{1, \ldots, m!\}$. Under anonymity $f(p) = f(q)$, $p \in M_f \iff q \in M_f$ and $p \in B^k \iff q \in B^k$ for all such profiles q. Let

$$l := \begin{cases} m + 1 & \text{if } m = 4 \\ m + 2 & \text{if } m \geq 5. \end{cases}$$

Suppose that the assumptions of the theorem are satisfied, and that

$$|M_f| \leq n\left(\frac{m!}{3} - 1\right)(m - 2).$$

Step 1: $B^l \cap M_f = \emptyset$.

Suppose to the contrary that there is a $p \in B^l \cap M_f$. As $p \in B^l$, at least l of the $c(p, i), i \in \{1, \ldots, m!\}$, satisfy $c(p, i) \geq 1$. Hence, by Lemma 64

$$\prod_{i=1}^{m!} c(p, i)! \leq \left(\sum_{i=1}^{m!} c(p, i) - l\right)! = (n - l)!.$$

So,

$$|M_f| \geq \frac{n!}{\prod_{i=1}^{m'} c(p, i)!} \geq \frac{n!}{(n-l)!} = n\alpha(l),$$

where $\alpha(l) := \frac{(n-1)!}{(n-l)!}$.

Now, if $m = 4$, then as $n \geq m + 1$,

$$\alpha(l) = \prod_{j=1}^{l-1}(n - j) \geq \prod_{j=1}^{m}(m + 1 - j) = m!$$

$$= 3\frac{m!}{3} > 2(\frac{m!}{3} - 1) = (m - 2)(\frac{m!}{3} - 1),$$

and, if $m \geq 5$, then, as $n \geq m + 2$,

$$\alpha(l) = \prod_{j=1}^{l-1}(n - j) \geq \prod_{j=1}^{m+1}(m + 2 - j) = (m + 1)!$$

$$> (m + 1)(\frac{m!}{3} - 1) > (m - 2)(\frac{m!}{3} - 1).$$

This contradicts

$$|M_f| \leq n(m - 2)(\frac{m!}{3} - 1),$$

so step 1 is proven.

Step 2: There are $b, c \in A$, $b \neq c$, such that $f(B^2) \subseteq \{b, c\}$.

By step 1 f is strategy-proof on B^l. So, by Theorem 63, $|f(B^l)| \leq 2$. Let $b, c \in A$, $b \neq c$, be such that $f(B^l) \subseteq \{b, c\}$. Let k be the smallest number such that

$$f(B^k) \subseteq \{b, c\}.$$

Then $k \geq 2$ by surjectivity of f, and $k \leq l$ since $f(B^l) \subseteq \{b, c\}$. We show that $k = 2$. As $f(B^{k-1}) \nsubseteq \{b, c\}$, there is a $p \in B^{k-1} - B^k$, such that $f(p) = a \in A - \{b, c\}$. Furthermore, as $k \leq n$, and $p \in B^{k-1} - B^k$, there must be agents $i, j \in N, i \neq j$, such that $p(i) = p(j)$, which implies that $p(N) = p(N - \{i\})$. Let t be such that

$$\{(a, c), (b, c)\} \subset t \text{ and } t \notin p(N).$$

As $|p(N)| = k - 1$, there are at least $\frac{m!}{3} - (k - 1)$ such preferences t. Then $t \notin p(N)$ implies that $q = (p_{-i}, t) \in B^k$. So $f(q) \in \{b, c\}$, $f(p) = a$ and $(f(q), f(p)) \notin t$. Hence, any such q is manipulable. So, by anonymity

$$|M_f| \geq \left(\frac{m!}{3} - (k - 1)\right)\frac{n!}{(n - (k - 1))!}.$$

But, by Lemma 64 the term on the right hand side is greater than $n(m - 2)(\frac{m!}{3} - 1)$ if $k - 1 \geq 2$, contradicting $|M_f| \leq n(m - 2)(\frac{m!}{3} - 1)$. Therefore, $k = 2$. This proves step 2.

Let $a \in A$ be an alternative different from b and c as in step 2. We define

$$P^{a, \{b, c\}} := \{t \in P \mid \{(a, b), (a, c)\} \subseteq t\}.$$

Let $A = \{a_1, a_2, \ldots, a_{m-2}, b, c\}$. As $f(B^2) \subseteq \{b, c\}$, we have

$$f(P^N - B^2) \supseteq \{a_1, a_2, \ldots, a_{m-2}\}$$

by surjectivity. Hence, for any $a_j, j \in \{1, 2, \ldots, m-2\}$, there is a $t_j \in P$ such that $f(t_j^N) = a_j$. Let $p^j := t_j^N, j \in \{1, 2, \ldots, m-2\}$. Then for any $t \in P^{a_j,\{b,c\}} - \{t_j\}, k \in N$, $f(p^j_{-k}, t) \in \{b, c\}$ and $f(p^j) = a_j$, so $q = (p^j_{-k}, t)$ is manipulable for all $k \in N$. There are

$$n | P^{a_j,\{b,c\}} - \{t_j\} |$$

such q, as $n \geq 3$ these manipulable profiles q are different for all $j \in \{1, 2, \ldots, m-2\}$. So,

$$n(m-2)(\frac{m!}{3} - 1) \geq |M_f| \geq n \sum_{i=1}^{m-2} |P^{a_j,\{b,c\}} - \{t_j\}|. \tag{6.2}$$

Now, $|P^{a_j,\{b,c\}} - \{t_j\}| \geq \frac{m!}{3} - 1$, and equality holds if and only if $t_j \in P^{a_j,\{b,c\}}$. So, by inequality (6.2), $t_j \in P^{a_j,\{b,c\}}, j \in \{1, 2, \ldots, m-2\}$,

$$|M_f| = n(m-2)(\frac{m!}{3} - 1)$$

and

$$M_f = \{(p^j_{-k}, t) \mid k \in N, j \in \{1, 2, \ldots, m-2\}, t \in P^{a_j,\{b,c\}} - \{t_j\}\}.$$

Step 3: Let b, c be as in step 2. Then

$$f(P^N - \{p^1, p^2, \ldots, p^{m-2}\}) = \{b, c\}$$

Let $p \in P^N - \{p^1, p^2, \ldots, p^{m-2}\}$. Suppose that $f(p) \in \{a_1, a_2, \ldots, a_{m-2}\}$. Then $p \in P^N - B^2$ and for all $t \in P^{f(p),\{b,c\}} - \{p(1)\}$ we have $f(p_{-1}, t) \in f(B^2) \subseteq \{b, c\}$. So $(p_{-1}, t) \in M_f$, implying that $p \in \{p^1, p^2, \ldots, p^{m-2}\}$, as $n \geq 3$. With surjectivity this proves step 3.

Step 4: Let $p \notin \{p^1, p^2, \ldots, p^{m-2}\}$ and suppose that $(b, c) \in t_j$ for some $j \in \{1, 2, \ldots, m-2\}$. Then,

$$f(p) = b \text{ if and only if } (b, c) \in p(i) \text{ for all } i \in N.$$

Let $t, \bar{t} \in P$ be such that $t = \ldots b \ldots a_j \ldots c \ldots$, $\bar{t} = \ldots c \ldots a_j \ldots b \ldots$. Let $k \in N$. As $t, \bar{t} \notin P^{a_j,\{b,c\}}$, f is strategy-proof at $(p^j_{-k}, t) \in B^2$ and $(p^j_{-k}, \bar{t}) \in B^2$. This implies that $f(p^j_{-k}, t) = b$ and $f(p^j_{-k}, \bar{t}) = c$.

Suppose that there is a $i \in N$ such that $(c, b) \in p(i)$, without loss of generality $i = n$. Let

$$r^u(l) := \begin{cases} p(l) & \text{if } l \leq u \\ (p^j_{-n}, \bar{t})(l) & \text{if } l > u, \end{cases}$$

$u \in \{0, \ldots, n\}$. Then $r^0 = (p^j_{-n}, \bar{t})$, $r^n = p$ and as $\bar{t} \notin P^{a_j,\{b,c\}}$,

$$r^u \notin M_f \cup \{p^1, p^2, \ldots, p^{m-2}\}, u \in \{0, \ldots, n-1\}.$$

Suppose that $f(r^u) = c$, $u \in \{0, \ldots, n-2\}$, then by $r^u \notin M_f$

$$(f(r^u), f(r^{u+1})) \in r^u(u+1) = t_j.$$

As $(b, c) \in t_j$, $f(r^{u+1}) \in \{b, c\}$ and $f(r^u) = c$ it follows that $f(r^{u+1}) = c$. As $f(r^0) = f(p^j_{-n}, \bar{t}) = c$ we obtain $f(r^{n-1}) = f(p) = c$ by induction. By strategy-proofness at $p = r^n = (r^{n-1}_{-n}, p(n))$, $(c, b) \in p(n)$, and $f(p) \in \{b, c\}$ it follows that $f(r^n) = c$. This shows that

$$f(p) = b \text{ only if } (b, c) \in p(i) \text{ for all } i \in N.$$

Suppose that $(b, c) \in p(i)$ for all $i \in N$, so in particular $(b, c) \in p(1)$. Let

$$r^u(l) := \begin{cases} (p^j_{-1}, t)(l) & \text{if } l \leq u \\ p(l) & \text{if } l > u, \end{cases}$$

$u \in \{0, \ldots, n\}$. Then $r^0 = p$, $r^n = (p^j_{-1}, t)$, and as $t \notin P^{a_j, \{b, c\}}$,

$$r^u \notin M_f \cup \{p^1, p^2, \ldots, p^{m-2}\}, u \in \{1, \ldots, n\}.$$

As $r^n = (p^j_{-1}, t)$, $f(r^n) = b$. Let $u \in \{1, 2, \ldots, n\}$ and suppose that $f(r^u) = b$. Then by $r^u \notin M_f$ $(f(r^u), f(r^{u-1})) \in r^u(u) = (p^j_{-1}, t)(u) \in \{t, t_j\}$. As $(b, c) \in t$, $(b, c) \in t_j$, $f(r^{u-1}) \in \{b, c\}$ and $f(r^u) = b$ it follows that $f(r^{u-1}) = b$. So, $f(r^n) = f(p_{-1}, t) = b$ implies by induction that $f(r^0) = f(p) = b$. This shows that

$$f(p) = b \text{ if } (b, c) \in p(i) \text{ for all } i \in N,$$

and we have proven step 4.

Without loss of generality $(b, c) \in t^1$. Then step 4 and the well-definedness of f imply that $(b, c) \in t_j$ for all $j \in \{1, 2, \ldots, m-2\}$. This completes the proof. ∎

Almost dictatorial social choice functions have $(n-1)(\frac{m!}{2} - 1) + 1$ manipulable profiles, see Maus et al. (2004c, see Chapter 5) for a definition of almost dictatorial social choice functions and a proof of this statement. As, if $m \geq 4$,

$$n(m-2)(\frac{m!}{3} - 1) > n(\frac{m!}{2} - 1)$$
$$= (n-1)(\frac{m!}{2} - 1) + (\frac{m!}{2} - 1)$$
$$> (n-1)(\frac{m!}{2} - 1) + 1,$$

the minimally manipulable anonymous social choice functions have strictly more manipulable profiles than one has to admit when demanding only nondictatoriality. This is in contrast to the three alternative case where we have, for $m = 3$,

$$n(m-2)(\frac{m!}{3} - 1) = n < 2n - 1 = (n-1)(\frac{m!}{2} - 1) + 1,$$

so the anonymous social choice functions outperform the almost dictatorial social choice functions in terms of manipulability. We conclude this section with some proportions of manipulable profiles of social choice functions to give an impression in relative terms of the manipulability that has to be admitted. The following table contains the proportion of profiles that are manipulable for the social choice functions characterized in Theorem 65, the almost dictatorial social choice functions, Plurality and Borda rule. The numbers for the last two are taken from Aleskerov and Kurbanov (1999), where tie-breaking according to a fixed order of alternatives is used to decide on ties. We restrict the tables to three and four alternatives. Loosely speaking, these tables indicate that there is still a lot to discover in the space between classical social choice functions and social choice functions obtained by minimizing manipulability. To avoid wrong impressions we note that Slinko (2002) has shown that the proportions of manipulable profiles of Plurality and Borda rule go to zero at a speed of at least $O(\frac{1}{\sqrt[2]{n}})$.

(m, n)	Almost dictatorial	Theorem 65	Plurality	Borda
$(3, 2)$	8,333333%	5,555556%	11,11%	38,89%
$(3, 3)$	2,314815%	1,388889%	16,67%	23,61%
$(3, 4)$	0,540123%	0,308642%	18,52%	31,02%
$(3, 5)$	0,115741%	0,064300%	23,15%	28,55%
$(3, 6)$	0,023577%	0,012860%	23,93%	27,82%
$(3, 7)$	0,004644%	0,002501%	25,73%	27,05%
$(3, 8)$	0,000893%	0,000476%	27,39%	25,99%
$(3, 9)$	0,000169%	0,000089%	27,44%	24,98%
$(3, 10)$	0,000031%	0,000017%	28,55%	24,06%

(m, n)	Almost dictatorial	Theorem 65	Plurality	Borda
$(4, 2)$	0,173611111111%	4,861111111111%	20,83%	60,42%
$(4, 3)$	0,014467592593%	0,303819444444%	29,43%	51,22%
$(4, 4)$	0,000904224537%	0,016878858025%	32,47%	50,02%
$(4, 5)$	0,000050234697%	0,000879107189%	37,38%	50,44%
$(4, 6)$	0,000002616390%	0,000043955359%	38,91%	47,90%
$(4, 7)$	0,000000130820%	0,000002136719%	40,55%	46,43%
$(4, 8)$	0,000000006359%	0,000000101749%	41,33%	44,85%
$(4, 9)$	0,000000000303%	0,000000004769%	41,99%	43,15%
$(4, 10)$	0,000000000014%	0,000000000221%	41,95%	41,78%

6.8   Conclusion

We show that a Gibbard-Satterthwaite like result holds also on sets of profiles with a certain diversity. This is used to characterize the minimally manipulable anonymous social choice functions for more than three alternatives. It turns out that these generalize strategy-proof

two alternative imputation status-quo voting, and the minimally manipulable three alternative nondictatorial social choice functions. However, contrary to the three alternative case they are not less manipulable than the almost dictatorial social choice functions. So, they do not constitute the set of minimally manipulable nondictatorial social choice functions. Moreover, the way in which they achieve minimal manipulability is at the expense of treating alternatives unequally. This suggests that a property ensuring a more equal treatment of alternatives should be added. The natural candidate is neutrality, but unfortunately anonymity and neutrality exclude each other whenever the number of alternatives can be written as a sum of non-trivial divisors of the number of agents, see Moulin (1983, p. 25). Nevertheless, this will be an issue for further research, either demanding weaker axioms than neutrality to ensure a more equal treatment of alternatives, or weakening anonymity and demanding neutrality. Also other principles of voting could be added as properties, such as respecting unanimity, Pareto-optimality, absolute plurality, plurality or Condorcet-winners.

Chapter 7

Unanimity and Anonymity

This chapter is concerned with the minimal number of profiles at which a unanimous and anonymous social choice function is manipulable. The lower bound is derived when there are three alternatives to choose from. Examples of social choice functions attaining the lower bound are given. We conjecture that these examples are in fact all minimally manipulable social choice functions. Since some of these examples are even Pareto optimal, we have also derived the lower bound for Pareto optimal and anonymous social choice functions. Some of the minimally manipulable Pareto optimal and anonymous social choice functions can be interpreted as status quo voting.[1]

7.1 Introduction

A well-known result of Gibbard (1973) and Satterthwaite (1975) states that any strategy-proof surjective social choice function on more than two alternatives is dictatorial. This implies that anonymity and unanimity are not compatible with strategy-proofness. Here we investigate how incompatible these are, i.e. how much manipulability we have to allow at anonymous and unanimous social choice functions.

There are a few studies on the degree of vulnerability for strategic behaviour of classical social choice functions, such as Borda, Plurality, etc.. For example, Aleskerov and Kurbanov(1999) contains simulation and enumeration results on 26 different social choice functions for different indices of manipulability. Slinko (2002) counts the number of instabile profiles of classical social choice functions, which is an upper bound for the number of manipulable profiles of these social choice functions. On the other hand, little is known about to what extent manipulability has to be admitted by social choice functions satisfying certain properties. Results on this may provide lower bounds to which then all social choice functions satisfying these properties can be compared. A first investigation was pioneered by Kelly (1988), who found the minimal number of manipulable profiles for nondictatorial surjective social choice functions with three alternatives and two agents

---

and also formulated several conjectures. This line of research was continued by Fristrup and Keiding (1998), who determined the minimal number of manipulable profiles for two agents and any number of alternatives. Maus et al. (2004c) consider the three alternative case and show that for any number of agents larger than two there are six minimally manipulable nondictatorial surjective social choice functions, and these are even anonymous. However, they also consider minimally manipulable unanimous and nondictatorial social choice functions. These turn out to be nondictatorial only in a minimal sense, namely at only one profile.

Here we will exclude such social choice functions by demanding that agents are treated anonymously and choice is unanimous. Maus et al. (2004b) contain results for this case if unanimity is replaced with surjectivity. In particular, anonymous social choice functions are nondictatorial. So, we are interested in the minimal number of manipulable profiles that an anonymous unanimous social choice function has to admit. It is well known that in case of two alternatives any monotonic social choice function is strategy-proof. As monotonicity and anonymity are compatible (see e.g. May (1952)) in case of two alternatives this minimal number is zero. We consider the three alternatives case and show that the minimal number is $2^n - 2$. Furthermore, we give examples of social choice functions reaching that lower bound and point out that some of these examples are even Pareto optimal. So, $2^n - 2$ is also the lower bound for Pareto optimal and anonymous social choice functions. After that we argue that status quo voting is an example of a Pareto optimal and anonymous social choice function that is minimally manipulable for three alternatives.

It should be noted that the way in which we measure manipulability, namely by counting manipulable profiles, is not the only possible one. In fact there are a lot of variations of this, where for example profiles are counted by the number of agents that can manipulate them, or the severity of the manipulation is taken into account. The severity is measured by the distance in the agent's ranking between what he achieves with and without manipulation. Aleskerov and Kurbanov (1999) contains an overview of such variations and numerical results on the manipulability of 26 social choice functions according to these measurements.

The chapter is organized as follows. Section 7.2 contains the basic definitions. Then, Section 7.3 shows some special results about the manipulability of two and three agent social choice functions that will be used in our proof of the main result. The main result, saying that any unanimous and anonymous social choice function over three alternatives has at least $2^n - 2$ manipulable profiles, where n is the number of agents, is proven in Section 7.4. After that we provide examples of social choice functions attaining that lower bound in Section 7.5. In Section 7.6 we see that some of the examples from Section 7.5 are even Pareto optimal. Pareto optimality is a stronger requirement than unanimity. This shows, given anonymity, that though Pareto optimality is stronger it does not raise the number of manipulable profiles one has to admit compared to imposing unanimity. Finally, we show that the minimally manipulable Pareto optimal and anonymous social choice functions have a nice interpretation as status quo voting rules.

7.2   Preliminaries

We denote the cardinality of a set S by $|S|$. Let $\mathbb{N}$ be the set of natural numbers, and let $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$. Let $\lceil \; \rceil$ denote the upper entier. So for $n \in \mathbb{N}_0$ $\left\lceil \frac{n}{2} \right\rceil := \min\{l \in \mathbb{N}_0 \mid l \geq \frac{n}{2}\}$.

Let A be a finite set of alternatives, $m := |A| \geq 3$, and $N = \{1, \ldots, n\}$ be a finite set of agents, $n \geq 2$.

Let $t \subset A \times A$. We call t *complete* if for all $x, y \in A$ $(x, y) \in t$ or $(y, x) \in t$. Note that completeness of t implies $(x, x) \in t$ for all $x \in A$. We call t *transitive* if for all $x, y, z \in A$ $(x, y) \in t$ and $(y, z) \in t$ implies $(x, z) \in A$. We call t *antisymmetric* if for all $x, y \in A$ $(x, y) \in t$ and $(y, x) \in t$ implies that $x = y$.

A *preference* $t \subset A \times A$ is a *linear order* (complete, transitive, antisymmetric) on A. Let P denote the set of all preferences. Suppose that $A = \{x_1, x_2, \ldots x_m\}$. By completeness, transitivity and antisymmetry we can write conveniently

$$t = x_1 x_2 \ldots x_m$$

for the preference t such that $(x_i, x_j) \in t$ if and only if $i \geq j$, $i, j \in \{1, 2, \ldots, m\}$,

$$t = \ldots x \ldots y \ldots$$

if we want to express only that x is strictly preferred to y, and

$$t = x \ldots$$

if we want to express only that x is preferred to all other alternatives.

A *profile* is a map $p : N \to P$. Thus, a profile assigns to every agent i a preference $p(i)$ over the alternatives. For a nonempty subset S of N we denote by $p|_S$ the restriction of the map p to the domain S. The set of all profiles is denoted by P^N.

For a permutation σ of N and a profile $p \in P^N$ let $p \circ \sigma$ be the profile given by $(p \circ \sigma)(i) := p(\sigma(i))$ for all $i \in N$. Two profiles $p, q \in P^N$ are called *anonymously equivalent* if there is a permutation σ of N such that $q = p \circ \sigma$. Then $|p^{-1}(t)| = |q^{-1}(t)|$ for all $t \in P$. We denote anonymous equivalence between p and q in P^N by $p \sim q$ and write $[p] := \{q \in P^N \mid q \sim p\}$. For all $p \in P^N$,

$$\|[p]\| := \frac{n!}{\prod_{t \in P} |p^{-1}(t)|!}.$$

A *social choice function* is a function $f : P^N \to A$. Hence, a social choice function selects a unique alternative $f(p)$ at every profile p.

A social choice function is called *surjective* if any alternative in A is chosen at least once, i.e. if $f(P^N) = A$. In literature this is also known as citizen-sovereignty. Moreover, a social choice function is called *unanimous* if $f(p) = x$ for all profiles $p \in P^N$ such that $p(i) = x \ldots$ for all $i \in N$. Clearly, unanimity is stronger than surjectivity. Throughout this chapter we assume that any social choice function is at least surjective.

A social choice function is called *anonymous* if $f([p]) = \{f(p)\}$ for all $p \in P^N$. Thus, anonymous social choice functions are symmetric in the arguments. In a sense they treat agents equally.

In contrast to anonymity, the following social choice functions respect only the preference of a single agent. Let $d \in N$ and for any profile p let

$$dict_d(p) := x$$

where x is such that $(x, y) \in p(d)$ for all $y \in A$. So x is the most preferred alternative of agent d in p. The social choice functions $f = dict_d$, $d \in N$, are called *dictatorial* social choice functions, and agent d is called a *dictator*. A social choice function f is called *nondictatorial* if there is no agent d such that $f = dict_d$.

We are interested in strategic behaviour of individuals when facing cooperative decision-making as captured by social choice functions. This is formalized by the following definitions. Let $f : P^N \to A$ be a social choice function. Let $p \in P^N$ be a profile. Then each profile q such that, for some $i \in N$, $q|_{N-\{i\}} = p|_{N-\{i\}}$ and $q(i) \neq p(i)$, is called an *i-devation* from p. Letting $t := q(i)$ we use the notation $q = (p_{-i}, t)$. If it is not important which agent deviates from p to q we call q a *deviation* from p. A profile p is called *manipulable* (under f) if there is an agent that is better off by being dishonest about his preference, i.e. if there is an $i \in N$, and an i-deviation q such that

$$(f(p), f(q)) \notin p(i).$$

In this case we say that p is *manipulable towards* q (under f). Let

$$M_f := \{p \in p^N \mid p \text{ is manipulable under } f\}.$$

Note that $p \in M_f$ implies that $[p] \subseteq M_f$. A social choice function is called *strategy-proof* if $M_f = \emptyset$, otherwise it is said to be *manipulable*. Let F be a nonempty set of social choice functions. We call f^* *minimally manipulable* if $f^* \in \arg\min_{f \in F} |M_f|$.

The prominence of the dictatorial rules arises from the following impossibility result due to Gibbard (1973) and Satterthwaite (1975).

Theorem 66 *Let A be a finite set of alternatives, $|A| \geq 3$. Let $f : P^N \to A$ be a nondictatorial surjective social choice function. Then*

$$|M_f| \geq 1.$$

However, this theorem makes no statement about the number of manipulable profiles such social choice functions admit. This question has been solved by Kelly (1988) for two agents and three alternatives, and by Fristrup and Keiding (1998) for two agents, and any number of alternatives larger than three. Maus et al. (2004c, see Chapter 5) consider the case of unanimous and nondictatorial social choice function for any numbers of agents and

alternatives larger than or equal to three, and the case of surjective and nondictatorial social choice functions for three alternatives. Maus et al. (2004b, see Chapter 6) consider the case of anonymous and surjective social choice functions, but have to leave some cases open.

Here we address the case of unanimous and anonymous social choice functions over three alternatives. So, let F be the set of all such social choice functions with a set of agents N such that $|N| = n \geq 2$. We will need the following simple lemma.

Lemma 67 *Let $k, l \in \mathbb{N}$. Then*

$$k! l! \leq (k + l - 1)!$$

Proof. $\frac{(k+l-1)!}{k! l!} = \frac{1}{k+l} \binom{k+l}{k} \geq 1$ as $k + l > k > 0$. ∎

7.3 Special results for two and three agents

Since the connection of the results of this section with the main result is far from obvious we sketch the proof of the main result beforehand.

Let $f \in F$. We will prove that there are natural numbers k_1, k_2 such that $\frac{n}{2} \leq k_1 \leq k_2 \leq n$, and profiles $q_i^k \in M_f$ for all $\frac{n}{2} \leq k \leq k_i - 1, k \in \mathbb{N}, i \in \{1, 2\}$ with the following properties. Each such q_i^k shows that there are at least $\binom{n}{k}$ manipulable profiles in M_f, which can be equal to the ones from another q_i^k only if $k = \frac{n}{2}$ for both profiles. Furthermore, we show that if $k_i < n, i \in \{1, 2\}$, then there is a $p_i \in M_f$ that yields at least another $\sum_{j=k_i}^{n-1} \binom{n}{j}$ different manipulable profiles in M_f. Altogether we will have that in any case

$$|M_f| \geq \sum_{j=1}^{n-1} \binom{n}{k} = 2^n - 2,$$

which is our main result. We find the q_i^k by applying Theorem 68 to two agent social choice functions $f_k, \frac{n}{2} \leq k < k_2$, defined from f. To show that the manipulable profiles arising from the same $f_k, k \neq \frac{n}{2}$, are different we need the statements of Lemma 69. To apply Theorem 68 and Lemma 69 we need that the $f_k, k < k_2$, are nondictatorial. This will be shown by applying Lemma 70 to three person social choice function g defined from f. This will also show that the p_i exist if $k_i < n$.

Now, we present the mentioned results for two and three agent social choice functions. Throughout the rest of this section let $N = \{1, 2\}$ if there are two agents, and $N = \{1, 2, 3\}$ if there are three agents. Two agent social choice functions can be depicted in 6x6 tables, since we have only six preferences. We let agent $1's$ preference be constant in rows and agent $2's$ preference constant in columns. The following result is taken from Kelly (1988). It can also be proven by considering 6x6 tables.

Theorem 68 *Let $f : P^N \to A$ be a surjective two agent social choice function. Then f is either dictatorial or $|M_f| \geq 2$.*

Let $D_f : M_f \to 2^{N \times P}$ be the correspondence assigning to each manipulable profile the set of agents that can manipulate and the preferences they can use to manipulate, i.e. for all $p \in M_f$,

$$D_f(p) := \{(i, t) \in N \times P \mid (f(p), f(p_{-i}, t)) \notin p(i)\}.$$

The first result shows some properties of D_f for unanimous and nondictatorial two agent social choice function.

Lemma 69 *Let $f : P^N \to A$ be a unanimous and nondictatorial two agent social choice function. Then for all $p, q \in M_f$ there are $(i, t_p) \in D_f(p), (j, t_q) \in D_f(q)$ such that $t_p \notin p(N), t_q \notin q(N)$. Furthermore we can choose i, j, t_p and t_q in such a way that, either*

(a) $i \neq j$,

(b) $p(N) \cup \{t_p\} \neq q(N) \cup \{t_q\}$, or,

(c) there is a labeling of the alternatives in A such that

$$p = (bca, cab), (i, cba) \in D_f(p)$$
$$q = (cba, bca), (i, cab) \in D_f(q)$$
$$f(p) = a.$$

Proof. By Theorem 66 we can find $p, q \in M_f$. First of all we show that for any $p \in M_f$ and $(k, t_p) \in D_f(p)$ such that $t_p \in p(N)$, there is a preference $\tilde{t}_p \notin p(N)$ such that $(k, \tilde{t}_p) \in D_f(p)$. This shows that we can choose $(i, t_p) \in D_f(p), (j, t_q) \in D_f(q)$ such that $t_p \notin p(N), t_q \notin q(N)$. Suppose that without loss of generality $k = 1$. So $(1, t_p) \in D_f(p)$ and $t_p \in p(N)$, hence $p = (p(1), t_p)$. Let $\tilde{t}_p \neq t_p$ be the preference such that $t_p = x \ldots$ and $\tilde{t}_p = x \ldots$. Then $f(\tilde{t}_p, t_p) = f(t_p, t_p)$ by unanimity of f. So $(1, \tilde{t}_p) \in D_f(p)$ as $(1, t_p) \in D_f(p)$, $\tilde{t}_p \notin p(N) = \{p(1), t_p\}$.

Note that in (a) $p = q$ is not excluded. Suppose that (a) and (b) are not true. Then all $p, q \in M_f$ are manipulable by the same agent i, without loss of generality let $i = 1$.

Let $O_2(t) := \{f(p) \mid p(1) = t\}$ be the set of all opportunities that agent 2 has when the preference of agent 1 is fixed to t. Let $A = \{x, y, z\}$. Consider $t = xyz$ in P. Then $x \in O_2(t)$ by unanimity. Note that fixing agent 1's preferences corresponds to looking at rows only in the tables describing f. Agent 2's inability to manipulate implies then that we can only have a certain type of rows.

	2 :	xyz	xzy	yxz	yzx	zxy	zyx
$1. O_2(xyz) = \{x\}$	$1 : xyz$	x	x	x	x	x	x
$2. O_2(xyz) = \{x, y\}$	$1 : xyz$	x	x	y	y	x	y
$3. O_2(xyz) = \{x, z\}$	$1 : xyz$	x	x	x	z	z	z
$4. O_2(xyz) = A$	$1 : xyz$	x	x	y	y	z	z

Similar tables hold for any $O_2(t), t \in P$.

Now, we fix $\{x, y, z\}$ such that $p(1) = xyz$. Remember that $p, q \in M_f$ are such that agent 1 manipulates at p and q and that

$$\{xyz, p(2), t_p\} = \{q(1), q(2), t_q\}.$$

Step 0: $x \in O_2(xyz)$.

This is a consequence of unanimity.

Step 1: $O_2(xyz) \neq \{x\}$.

If $O_2(xyz) = \{x\}$ then agent 1 cannot manipulate at p.

Step 2: $O_2(xzy) \neq \{x\}$.

Suppose contrapositive that $O_2(xzy) = \{x\}$. By step 1 there is an $u \in O_2(xyz) - \{x\}$. Let $t_1 \neq t_2$ be the two preferences such that $t_i = u \ldots, i = 1, 2$. Then $f(xyz, t_i) = u, i \in \{1, 2\}$, since agent 2 cannot manipulate, and so (xyz, t_i) is manipulable by agent 1 towards (xzy, t_i) for $i \in \{1, 2\}$. But then (b) holds, contradicting our assumptions. Hence, $O_2(xzy) \neq \{x\}$.

Step 3: $z \in O_2(xzy)$.

In view of step 2 suppose contrapositive that $O_2(xzy) = \{x, y\}$. Then $f(xzy, zyx) = y$, since agent 2 cannot manipulate, and $f(zxy, zyx) = z$ by unanimity. So $r = (xzy, zyx)$ is manipulable by agent 1 towards $(zxy, zyx) = (t_r, r(2))$. As $p(1) = xyz \notin r(N) \cup \{t_r\}$ this contradicts our assumption that (b) does not hold.

Step 4: $z \in O_2(xyz)$.

In view of step 1 suppose contrapositive that $O_2(xyz) = \{x, y\}$. Then $f(xzy, zxy) = z$ and $f(xyz, zxy) = x$, since agent 2 cannot manipulate. Hence, $r = (xzy, zxy)$ is manipulable by agent 1 towards $(xyz, zxy) = (t_r, r(2))$. As (b) does not hold we have $p(N) \cup \{t_p\} = \{xyz, p(2), t_p\} = r(N) \cup \{t_r\} = \{xyz, xzy, zxy\}$. If $p(2) \in \{xyz, xzy\}$ then $f(p) = x$ by unanimity. If $p(2) = zxy$, then $f(p) = x$ because $O_2(xyz) = \{x, y\}$ and agent 2 cannot manipulate. But $f(p) = x$ contradicts the manipulability of p by agent 1. Hence, $z \in O_2(xyz)$.

Step 5: $z \in O_2(yxz)$ and $z \in O_2(yzx)$.

Suppose contrapositive that there is a $t \in \{yxz, yzx\}$ such that $z \notin O_2(t)$, so $O_2(t) \subseteq \{x, y\}$. Let $r^1 = (xyz, zxy), r^2 = (xyz, zyx)$. By step 4 $z \in O_2(xyz)$, and so $f(r^1) = f(r^2) = z$. But then $f(r^1_{-1}, t) \in \{x, y\}$ and $f(r^2_{-1}, t) \in \{x, y\}$, so r^1 and r^2 are manipulable towards (r^1_{-1}, t) and (r^2_{-1}, t) respectively. Clearly, this contradicts the assumption that (b) is not satisfied. This proves step 5.

Step 6: $O_2(xzy) = \{x, z\}$.

Suppose contrapositive that in view of step 0 and 3 $O_2(xzy) = A$. By step 0 and 2 either $O_2(xyz) = \{x, z\}$ or $O_2(xyz) = A$. Consider first the case in which $O_2(xyz) = \{x, z\}$. Let $r^1 = (xzy, yxz), r^2 = (xzy, yzx)$. Then $f(r^1) = f(r^2) = y$, whereas $f(r^1_{-1}, xyz) = x$ and $f(r^2_{-1}, xyz) = z$. Hence, r^1 and r^2 are manipulable by agent 1 towards $(r^1_{-1}, xyz), (r^2_{-1}, xyz)$ respectively. But then (b) is satisfied a contradiction. So, suppose that $O_2(xyz) = A$. We know that p is manipulable by agent 1 and that $p(1) = xyz$. By steps 3 to 5, unanimity and

because agent 2 cannot manipulate $f(P \times \{zxy, zyx\}) = \{z\}$, hence no manipulation by agent 1 can occur if $p(2) \in \{zxy, zyx\}$. By unanimity agent 1 has no incentive to manipulate if $p(2) \in \{xyz, xzy\}$. Hence, $p(2) \in \{yxz, yzx\}$. Then $f(p) = y$ and $f(t_p, p(2)) = x$, since agent 1 manipulates p towards $(t_p, p(2))$. As $O_2(xzy) = A$ implies that $f(xzy, p(2)) = y$ (agent 2 cannot manipulate), and unanimity implies that $f(yxz, p(2)) = f(yzx, p(2)) = y$ we obtain $t_p \in \{zxy, zyx\}$. As $f(t_p, p(2)) = x$ the profile $r = (xzy, p(2)), f(r) = y$ ($O_2(xyz) = A$ and agent 2 cannot manipulate), is also manipulable by agent 1. Clearly, then with p and r (b) is satisfied, a contradiction. Hence, $O_2(xzy) = \{x, z\}$.

Step 7: $y \notin O_2(zyx)$.

Suppose contrapositive that $y \in O_2(zyx)$. Then $r = (zyx, yzx), f(r) = y$ (agent 2 cannot manipulate), is manipulable towards $(xzy, yzx) = (t_r, r(2))$, as $f(xzy, yzx) = z$ by step 6 and since agent 2 cannot manipulate. As $p(1) = xyz \notin r(N) \cup \{t_r\}$ this contradicts the assumption that (b) does not hold.

Step 8: $O_2(zyx) = \{z\}$

By step 7 $y \notin O_2(zyx)$. So suppose contrapositive that $O_2(zyx) = \{x, z\}$. Then $r = (zyx, yxz), f(r) = x$ (agent 2 cannot manipulate), is manipulable towards $(yzx, yxz) = (t_r, r(2))$, as $f(yzx, yxz) = y$ by unanimity. As $p(1) = xyz \notin r(N) \cup \{t_r\}$ this contradicts the assumption that (b) does not hold.

Step 9: $O_2(zxy) = \{z\}$.

Suppose contrapositive that there is an $t \in P$ such that $f(zxy, t) \neq z$. Because of unanimity $t \notin \{zxy, zyx\}$. Then $r = (zxy, t)$ is manipulable towards $(zyx, t) = (t_r, r(2))$ by step 8. As $p(1) = xyz \notin r(N) \cup \{t_r\}$ this contradicts the assumption that (b) does not hold unless $p(1) = t = xyz$ and $\{p(2), t_p\} = \{zxy, zyx\}$. But in that case $f(p_{-1}, t_p) = f(t_p, p(2)) = z$ by unanimity, but agent 1 in $p = (xyz, p(2))$ has no incentive to manipulate to a profile (p_{-1}, t_p) where outcome z obtains. So, this contradicts the assumption that agent 1 manipulates at p towards (p_{-1}, t_p). This finishes the proof of step 9.

Step 10: $O_2(yzx) = \{y, z\}$.

In view of step 5 suppose contrapositive that $O_2(yzx) = A$. Then $r = (yzx, zxy), f(r) = x$ (agent 2 cannot manipulate), is manipulable towards $(zxy, zxy) = (t_r, r(2))$ by step 9. As $p(1) = xyz \notin r(N) \cup \{t_r\}$ this contradicts the assumption that (b) does not hold.

We summarize now what we know about f by steps 1 to 10 in the following 6x6 table.

2 :	xyz	xzy	yxz	yzx	zxy	zyx
1 : xyz	x	x			z	z
1 : xzy	x	x	x	z	z	z
1 : yxz			y	y	z	z
1 : yzx	y	z	y	y	z	z
1 : zxy	z	z	z	z	z	z
1 : zyx	z	z	z	z	z	z

We finish the proof by considering the two cases for $O_2(xyz) \in \{\{x, z\}, A\}$.

If $O_2(xyz) = \{x, z\}$ then $r = (xyz, yzx), f(r) = z$, is manipulable towards $(yxz, yzx) =$

$(t_r, r(2))$, and r is the only manipulable profile such that $r(1) = xyz$. Hence, $p = r$. If (b) is not satisfied we must have $q(N) \cup \{t_q\} = \{xyz, yxz, yzx\}$. Since $q(1) \neq xyz$ we have $q(1) \in \{yxz, yzx\}$. Since q has to be manipulable this leads to $q = (yxz, xyz), t_q = yzx$. But then (c) holds, letting $i = 1$, $x = b$, $y = c$ and $z = a$.

On the other hand if $O_2(xyz) = A$, then $r = (xyz, yxz)$, $f(r) = y$, is manipulable towards $(xzy, yxz) = (t_r, r(2))$, and r is the only manipulable profile such that $r(1) = xyz$. Hence, $p = r$. If (b) is not satisfied we must have $q(N) \cup \{t_q\} = \{xyz, xzy, yxz\}$. Since $q(1) \neq xyz$ and clearly also $q(1) \neq xzy$ (otherwise $q \in \{(xzy, xyz), (xzy, yxz)\} \implies f(q) = x$) we have $q(1) = yxz$. So, $q \in \{(yxz, xyz), (yxz, xzy)\}$. Now, if $q = (yxz, xyz)$ then $t_q = xzy$, so agent 1 manipulates q towards (xzy, xyz), where he obtains the outcome $x = f(xzy, xyz)$. This implies that x is preferred to $f(q)$ under the preference $q(1) = yxz$, hence $f(q) = z$. This is a contradiction, since then agent 2 can manipulate q as well. So, only the case $q = (yxz, xzy), t_q = xyz$, remains. For the same reasons as in the other case the manipulability of q towards (xyz, xzy) by agent 1 implies that $f(q) = z$. But then (c) holds, letting $i = 1$, $x = c$, $y = b$, $z = a$, and exchanging the roles of p and q.

This finishes the proof of the lemma. ∎

Furthermore, we need the following result about the manipulability of unanimous three agent social choice functions where coalitions are dictators. Let $S \subseteq N$ be a nonempty coalition of agents. We call S a *dictator* if for all $t = xyz \in P$ and $q_{-S} \in P^{N-S}$ we have

$$f(t^S, q_{-S}) = x.$$

In this case we call f S−*dictatorial.*

Lemma 70 *Let $f : P^N \to A$ be a unanimous three agent social choice function. Suppose that f is $\{1, 2\}$−dictatorial and $\{1, 3\}$−dictatorial. Then, one of the following is true.*

(a) *Agent 1 is a dictator.*

(b) *There are anonymously inequivalent $p, q \in M_f$, manipulable towards $(p_{-i}, t_p), (q_{-j}, t_q), i, j \in \{1, 2, 3\}$, respectively, such that $|p(N)| = |q(N)| = 3$ and $p(N) \cup \{t_p\} \neq q(N) \cup \{t_q\}$.*

Proof. Let $f_t^2 : P^{\{1,2\}} \to A$, $t \in P$, be the two agent social choice function given by $f_t^2(p) := f(p(1), p(2), t)$ for all $p \in P^{\{1,2\}}$. Let $f_t^3 : P^{\{1,3\}} \to A$, $t \in P$, be the two agent social choice function given by $f_t^3(p) := f(p(1), t, p(3))$ for all $p \in P^{\{1,3\}}$. Suppose that (b) is not true. We show that then all f_t^2 are dictatorial with dictator 1, so (a) holds. By unanimity and $\{1, 3\}$−dictatoriality ($\{1, 2\}$−dictatoriality) of f we have some information about f_t^2 and f_t^3. We summarize this information in the following table for $t = xyz$, where agent 1's preference is as usually described by the rows. The other agent can be either agent 2 or agent 3, and the remaining of the three agents is fixed to the preference xyz.

2 (3) :	xyz	xzy	yxz	yzx	zxy	zyx
1 : xyz	x	x	x	x	x	x
1 : xzy	x	x				
1 : yxz			y			
1 : yzx				y		
1 : zxy					z	
1 : zyx						z

Let $r \in P^{\{1,2\}}$. Suppose that (b) is false. For steps 1 to 4 we omit the superscript $k = 2$.

Step 1: The second row can be filled with x, i.e.

$$f_{xyz}(xzy, t) = x \text{ for all } t \in P.$$

Suppose contrapositive that $u := f_{xyz}(xzy, t) \in \{y, z\}$ for some $t \in P$. Let $A = \{u, v, w\}$, and $p = (xzy, uvw, xyz), q = (xzy, uwv, xyz)$. If $f(p) \neq u$ then p is manipulable towards $(p_{-2}, t) = (p_{-2}, t_p)$, and if $f(p) = u$ then p is manipulable towards $(p_{-1}, xyz) = (p_{-1}, t_p)$. The same holds for q. Clearly, p and q are anonymously inequivalent and satisfy $|p(N)| = |q(N)| = 3$. Furthermore, $p(N) \cup \{t_p\} = q(N) \cup \{t_q\}$ implies that $t_p = uwv$ and $t_q = uvw$. As $u \neq x$ this implies that agent 2 is manipulating at p and q and so $t_p = t = t_q$, contradicting $t_p = uwv, t_q = uvw$. Hence, (b) is satisfied, a contradiction.

Step 2: f_{xyz} is unanimous.

In view of step 1 it remains to be shown that $f_{xyz}(yxz, yzx) = f_{xyz}(yzx, yxz) = y$ and $f_{xyz}(zxy, zyx) = f_{xyz}(zyx, zxy) = z$. But for example by applying step 1 to f_{yxz}^3 we obtain $f_{xyz}(yzx, yxz) = f_{yxz}^3(yzx, xyz) = y$. Likewise the other parts of unanimity of f_{xyz} follow.

Step 3: In the third and fourth row z cannot be chosen, i.e.

$$f_{xyz}(r) \neq z \text{ if } r(1) \in \{yxz, yzx\}.$$

Suppose contrapositive that $r(1) \in \{yxz, yzx\}$ and $f_{xyz}(r) = z$. Let $p = (r(1), zxy, xyz)$ and $q = (r(1), zyx, xyz)$. If $f(p) \neq z$ then p is manipulable towards $(p_{-2}, r(2)) = (p_{-2}, t_p)$, and if $f(p) = z$ then p is manipulable towards $(p_{-3}, r(1)) = (p_{-3}, t_p)$, as by $\{1, 3\}$−dictatoriality of f we have $f(r(1), zxy, r(1)) = y$. The same holds for q. Clearly, p and q are anonymously inequivalent and satisfy $|p(N)| = |q(N)| = 3$. Furthermore, $p(N) \cup \{t_p\} = q(N) \cup \{t_q\}$ implies that $t_p = zyx$ and $t_q = zxy$. As $r(1) \in \{yxz, yzx\}$ this implies that agent 2 is manipulating at p and q and so $t_p = r(2) = t_q$, contradicting $t_p = zyx, t_q = zxy$. Hence, (b) is satisfied, a contradiction.

Step 4: In the cell (yzx, zxy), y has to be chosen, i.e.

$$f_{xyz}(yzx, zxy) = y.$$

Let $p = (yzx, zxy, xyz)$ and $q = (yzx, zxy, xzy)$. By step 3 $f(p) \in \{x, y\}$. Suppose contrapositive to the claim of step 4 that $f(p) = x$. Then p is manipulable towards

$(p_{-1}, zxy) = (p_{-1}, t_p)$ by $\{1,2\}$–dictatoriality of f. If $f(q) \neq x$ then q is manipulable towards $p = (q_{-3}, xyz) = (q_{-3}, t_q)$, and if $f(q) = x$ then q is manipulable towards (q_{-1}, zxy) by $\{1,2\}$–dictatoriality of f. Clearly, p and q are anonymously inequivalent and satisfy $|p(N)| = |q(N)| = 3$. As $xzy \in q(N)$ but $xzy \notin p(N) \cup \{t_p\}$ (b) is satisfied, a contradiction.

Step 5: In the fourth row y has to be chosen, i.e.

$$f_{xyz}(yzx, t) = y \text{ for all } t \in P.$$

We show step 5 for $k = 2$. The case $k = 3$ works in the same way. By step 3 we have that $f^2_{xyz}(yzx, t) \in \{x, y\}$ for all $t \in P$. Furthermore, $f^2_{xyz}(yzx, zyx) = f^3_{zyx}(yzx, xyz) \in \{x, y\} \cap \{y, z\} = \{y\}$ by step 3 applied to f^2_{xyz} and f^3_{zyx}. Hence, in view of step 4, we know that the claim is true if $t \in \{yxz, yzx, zxy, zyx\}$. Now, suppose contrapositive to the claim that $f^2_{xyz}(yzx, t) = x$ for some $t \in \{xyz, xzy\}$. Let $p = (yzx, zxy, xyz)$. Then p is manipulable towards $(p_{-2}, t) = (p_{-2}, t_p)$, as $f(p) = f^2_{xyz}(yzx, yxz) = y$ and $f(p_{-2}, t) = f^2_{xyz}(yzx, t) = x$. Let $q^1 = (yxz, zxy, xyz)$, $q^2 = (yzx, xzy, xyz)$ and $q^3 = (yxz, xzy, xyz)$. By step 3 $f(q^i) \in \{x, y\}$ for all $i \in \{1, 2, 3\}$. We show that for all possible combinations of $f(q^i) \in \{x, y\}, i \in \{1, 2, 3\}$, (b) is satisfied, contradicting our assumptions.

If $f(q^1) = x$ then q^1 is manipulable towards $(q^1_{-1}, yzx) = (q^1_{-1}, t_{q^1}) = (yzx, zxy, xyz)$, as $f(yzx, zxy, xyz) = y$. Clearly, p and q^1 are anonymously inequivalent and satisfy $|p(N)| = |q^1(N)| = 3$. As $q^1(1) = yxz \notin p(N) \cup \{t_p\} \subset \{yzx, zxy, xyz, xzy\}$, (b) is satisfied. Hence, we are done unless $f(q^1) = y$.

If $f(q^2) = y$ then q^2 is manipulable towards $(q^2_{-2}, t) = (q^2_{-2}, t_{q^2}) = (yzx, t, xyz)$, as $f(yzx, t, xyz) = x$. Clearly, p and q^2 are anonymously inequivalent and satisfy $|p(N)| = |q^2(N)| = 3$. As $p(2) = zxy \notin q^2(N) \cup \{t_{q^2}\} \subset \{yzx, xzy, xyz\}$, (b) is satisfied, a contradiction. Hence, we are done unless $f(q^2) = x$.

If $f(q^3) = y$ then q^2 is manipulable towards $(q^2_{-1}, yxz) = (q^2_{-1}, t_{q^2}) = q^3$, as $f(q^2) = x$. Still, p and q^2 are anonymously inequivalent, satisfy $|p(N)| = |q^2(N)| = 3$, and $p(2) = zxy \notin q^2(N) \cup \{t_{q^2}\} \subset \{yzx, xzy, xyz, yxz\}$, so (b) is satisfied. Hence, $f(q^3) = x$. But then q^1 is manipulable towards $(q^1_{-2}, zxy) = (q^1_{-2}, t_{q^1}) = q^3$, as $f(q^1) = y$. Still, p and q^1 are anonymously inequivalent, satisfy $|p(N)| = |q^2(N)| = 3$, and $q^1(1) = yxz \notin p(N) \cup \{t_p\} \subset \{yzx, zxy, xyz, xzy\}$, so (b) is satisfied. This finishes the proof of step 4.

Step 6: In the third row y has to be chosen, i.e.

$$f_{xyz}(yxz, t) = y \text{ for all } t \in P.$$

We show step 5 for $k = 2$. The case $k = 3$ works in the same way. By step 3 we have that $f^2_{xyz}(yxz, t) \in \{x, y\}$ for all $t \in P$. Furthermore, $f^2_{xyz}(yxz, zyx) = f^3_{zyx}(yxz, xyz) \in \{x, y\} \cap \{y, z\} = \{y\}$ by step 3 applied to f^2_{xyz} and f^3_{zyx}. Hence, we know that the claim is true if $t \in \{yxz, yzx, zyx\}$. We extend this to $\{xzy, yxz, yzx, zxy, zyx\}$. So suppose contrapositive that $t \in \{xzy, zxy\}$ and $f^2_{xyz}(yxz, t) = x$. Then $p = (yxz, t, xyz)$ is manipulable by agent 1 towards $(yzx, t, xyz) = (t_p, t, xyz)$. As $t \in \{xzy, zxy\}$, $|p(N)| = 3$. Consider $\widetilde{t} \in \{xzy, zxy\} - \{t\}$ and let $q = (yxz, \widetilde{t}, xyz)$. If $f(q) = x$ then q is manipulable by agent

1 towards $(yzx, \tilde{t}, xyz) = (t_q, \tilde{t}, xyz)$. If $f(q) = y$ then q is manipulable by agent 2 towards $(yxz, t, xyz) = (yxz, t_q, xyz)$. In both cases $|q(N)| = 3$ as $\tilde{t} \in \{xzy, zxy\}$. Furthermore, $p(N) \cup \{t_p\} \neq q(N) \cup \{t_q\}$, as $\tilde{t} \in q(N)$ but $\tilde{t} \notin p(N) \cup \{t_p\} = \{yxz, yzx, t, xyz\}$. Hence, (b) is satisfied, a contradiction. So, $f_{xyz}(yxz, t) = y$ for all $t \in P - \{xyz\}$. It remains to be shown that $f_{xyz}(yxz, xyz) = y$. Suppose contrapositive that $f_{xyz}(yxz, xyz) = x$. Then $p = (yxz, zxy, xyz)$ and $q = (yxz, zxy, xyz)$, $f(p) = f(q) = y$, are both manipulable by agent 2 towards (yxz, xyz, xyz). So (b) is satisfied, a contradiction. This proves step 6.

Step 7: f_{xyz}^2 can be described by the following 6x6 table.

2 :	xyz	xzy	yxz	yzx	zxy	zyx
1 : xyz	x	x	x	x	x	x
1 : xzy	x	x	x	x	x	x
1 : yxz	y	y	y	y	y	y
1 : yzx	y	y	y	y	y	y
1 : zxy		z		z	z	z
1 : zyx		z		z	z	z

The first four rows and the $\{zxy, zyx\}^{\{1,2\}}$–cells follow by steps 1,5 and 6 applied to f_{xyz}^2. By steps 5 and 6 applied to f_{yzx}^3 we see that $f_{xyz}^2(t, yzx) = f_{yzx}^3(t, xyz) = z$ for all $t \in \{zxy, zyx\}$. Likewise, $f_{xyz}^2(t, xzy) = z$ for all $t \in \{zxy, zyx\}$ by applying steps 5 and 6 to f_{xzy}^3.

We prove now that f is dictatorial with dictator 1. This is equivalent to dictatoriality of all $f_t^2, t \in P$. To the contrary suppose that for some $t \in P$ f_t^2 is nondictatorial. Without loss of generality $t \in \{xyz, yxz\}$. Note that for both of these t the table of f_t^2 is similar. In view of step 7 this implies that there is a $r \in V := \{zxy, zyx\} \times \{xyz, yxz\}$, such that $u = f_t^2(r) \in \{x, y\}$. Without loss of generality $r(1) = zxy$, otherwise exchange the roles of x and y which is allowed as we only fixed $t \in \{xyz, yxz\}$. Let $A = \{u, v, z\}, p = (r(1), uzv, t)$. Then p is manipulable towards $(p_{-2}, r(2)) = (p_{-2}, t_p) = (r, t)$, as $f_t^2(r) = u$ and $f(p) = z$. If there is an $\tilde{r} \in \{zyx\} \times \{xyz, yxz\}$ such that $\tilde{u} = f_t^2(\tilde{r}) \in \{x, y\}$, $A = \{\tilde{u}, z, \tilde{v}\}$ then, likewise, one obtains that $q = (\tilde{r}(1), \tilde{u}z\tilde{v}, t))$ is manipulable towards $(q_{-2}, \tilde{r}(2)) = (q_{-2}, t_q) = (\tilde{r}, t)$. Clearly, such p, q, t_p and t_q satisfy (b), a contradiction. Hence, we are done unless

$$f_t^2(zyx, \tilde{t}) = z \text{ for all } \tilde{t} \in P.$$

Let $q = (zxy, yxz, t)$. If $f(q) = z$ then q is manipulable towards $(q_{-2}, r(2)) = (q_{-2}, t_q) = (r, t)$, as $f(r, t) \in \{x, y\}$. Clearly, p and q are anonymously inequivalent and satisfy $|p(N)| = |q(N)| = 3$. As $p(2) = uzv \notin q(N) \cup \{t_q\} \subset \{zxy, yxz, xyz\}$ (b) is satisfied, a contradiction. Hence, we are done unless $f(q) \in \{x, y\}$. But then q is manipulable towards $(q_{-1}, zyx) = (q_{-1}, t_q)$, as $f(zyx, q) = f_t^2(zyx, q(2)) = z$. Still, p and q are anonymously inequivalent, satisfy $|p(N)| = |q(N)| = 3$, and $p(2) = uzv \notin q(N) \cup \{t_q\} \subset \{zxy, yxz, xyz, zyx\}$. Hence, (b) is satisfied, and we have obtained contradictions in all cases. So, there is no

nondictatorial f_t and this finishes the proof. ∎

7.4   The lower bound $2^n - 2$

In this section we will prove $\min_{f \in F} |M_f| \geq 2^n - 2$, using the results from Section 7.3.

Let $k : P^N \to \{\lceil \frac{n}{2} \rceil, \ldots, n\}$ be given by

$$k(p) := \min\{|p_{-1}(T)| \mid T \subseteq P, |p^{-1}(T)| \geq \frac{n}{2}\}$$

and $l : P^N \to \{0, \ldots, \lceil \frac{n}{2} \rceil\}$ by $l(p) := n - k(p)$. Then the following lemma holds for profiles that contain at least three different preferences.

Lemma 71 *Let p be a profile such that $|p(N)| \geq 3$. Then*

$$\|p\| \geq \sum_{j=0}^{l(p)-1} \binom{n}{k(p) + j}.$$

Proof. Let T be such that $|p^{-1}(T)| = k(p)$. Then

$$
\begin{aligned}
\|p\| &= \frac{n!}{\prod_{t \in P} |p^{-1}(t)|} \\
&= \frac{n!}{\prod_{t \in T} |p^{-1}(t)|! \prod_{t \in P-T} |p^{-1}(t)|!} \\
&\geq \frac{n!}{(k(p) - |T| + 1)!(l(p) - |P| + |T| + 1)!} \\
&= \frac{k(p)!}{(k(p) - |T| + 1)!} \frac{l(p)!}{(l(p) - |P| + |T| + 1)!} \binom{n}{k(p)} \\
&\geq l(p) \binom{n}{k(p)} \geq \sum_{j=0}^{l(p)-1} \binom{n}{k(p) + j}.
\end{aligned}
$$

The first inequality holds as

$$\prod_{t \in \tilde{T}} |p^{-1}(t)|! \leq (k(p) - |\tilde{T}| + 1)!$$

for all $\tilde{T} \subseteq P$ by applying Lemma 67 repeatedly. Furthermore, by $|p(N)| \geq 3$ either $|P - T| \geq 2$ and

$$\frac{l(p)!}{(l(p) - |P| + |T| + 1)!} \geq \frac{l(p)!}{(l(p) - 1)!} = l(p)$$

or $|T| \geq 2$ and

$$\frac{k(p)!}{(k(p) - |T| + 1)!} \geq \frac{k(p)!}{(k(p) - 1)!} = k(p) \geq l(p),$$

which justifies the second inequality. Finally the third inequality is valid as

$$\binom{n}{k(p)} \geq \binom{n}{k(p) + j}$$

for all $j \in \{0, \ldots, l(p) - 1\}$, since $k(p) \geq \frac{n}{2}$. ∎

Let $M_f^3 := \{p \in M_f \mid |p(N)| \geq 3\}$, and if $M_f^3 \neq \emptyset$ let

$$p_1 \in \arg \min_{p \in M_f^3} k(p), \quad k_1 := k(p_1), \quad l_1 := l(p_1),$$

and if $M_f^3 = \emptyset$ let $k_1 = k_2 = n$. Furthermore, if $k_1 < n$ and $M_f^3 - [p_1] \neq \emptyset$ let

$$p_2 \in \arg \min_{p \in M_f^3 - [p_1]} k(p), \quad k_2 := k(p_2), \quad l_2 := l(p_2),$$

and if $M_f^3 - [p_1] = \emptyset$ let $k_2 = n$.

The next lemma shows, using the results from Section 7.3, that for all $k \in \{\lceil \frac{n}{2} \rceil, \ldots, k_1 - 1\}$ we have two profiles $q_1^k, q_2^k \in M_f$ such that $|q_1^k(N)| = |q_2^k(N)| = 2$ and $k(q_1^k) = k(q_2^k) = k$. If $k \neq \frac{n}{2}$ we show that these two profiles are anonymously inequivalent. For all $k \in \{k_1, \ldots, k_2 - 1\}$ we show that there is one profile $q_1^k \in M_f$ such that $|q_1^k(N)| = 2$ and $k(q_1^k) = k$. Now note that profiles $p, q \in P^N$, such that $|p(N)| = |q(N)| = 2$ and $k(p) \neq k(q)$ are anonymously inequivalent. Altogether, this then shows that, if n is odd,

$$|M_f| \geq \left| \bigcup_{k=\lceil \frac{n}{2} \rceil}^{k_2-1} [q_1^k] \cup \bigcup_{k=\lceil \frac{n}{2} \rceil}^{k_1-1} [q_2^k] \cup [p_1] \cup [p_2] \right|$$

$$= \sum_{k=\lceil \frac{n}{2} \rceil}^{k_2-1} \binom{n}{k} + \sum_{k=\lceil \frac{n}{2} \rceil}^{k_1-1} \binom{n}{k} + |[p_1]| + |[p_2]|$$

$$\geq \sum_{k=\lceil \frac{n}{2} \rceil}^{k_2-1} \binom{n}{k} + \sum_{k=\lceil \frac{n}{2} \rceil}^{k_1-1} \binom{n}{k} + \sum_{j=0}^{l_1-1} \binom{n}{k_1 + j} + \sum_{j=0}^{l_2-1} \binom{n}{k_2 + j}$$

$$= \sum_{k=1}^{n-1} \binom{n}{k} = 2^n - 2,$$

where we use Lemma 71 for the second inequality. Likewise one shows that $|M_f| \geq 2^n - 2$ if n is even, the only difference is that the two profiles $q_1^{\frac{n}{2}}, q_2^{\frac{n}{2}}$ do not have to be anonymously inequivalent.

Lemma 72 *Let $f : P^N \to A$ be a unanimous and anonymous social choice function. Let k_1, k_2, p_1, p_2 be as defined above. Then for all $k \in \{\lceil \frac{n}{2} \rceil, \ldots, k_1 - 1\}$ there are two profiles $q_1^k, q_2^k \in M_f$ such that $|q_1^k(N)| = |q_2^k(N)| = 2$ and $k(q_1^k) = k(q_2^k) = k$. If $k \neq \frac{n}{2}$ these two profiles are anonymously inequivalent. For all $k \in \{k_1, \ldots, k_2 - 1\}$ there is a profile $q_1^k \in M_f$ such that $|q_1^k(N)| = 2$ and $k(q_1^k) = k$.*

Proof. We will apply Lemma 69 to the following two agent social choice functions to show the existence of the q_i^k.

Let $k \in K := \{\lceil \frac{n}{2} \rceil, \ldots, n\}$, and $l := n - k$. Define $f_k : P^{\{1,2\}} \to A$ by

$$f_k(p) := f(p(1)^k, p(2)^l)$$

for all $p \in P^{\{1,2\}}$, where $(p(1)^k, p(2)^l)$ denotes a profile q in P^N, such that $|q^{-1}(p(1))| = k$ and $|q^{-1}(p(2))| = l$. As f is anonymous f_k is well defined and as f is unanimous, f_k is unanimous.

To apply Lemma 69 to a $f_k, k \leq k_2$, we have to show that f_k is nondictatorial. Let $\underline{d}$ be the smallest $k \in K$ such that f_k is dictatorial,

$$\underline{d} := \min\{k \in K \mid f_k \text{ is dictatorial}\}.$$

Note that f_n is dictatorial with dictator 1, so $\underline{d} \leq n$. We show that $k_2 \leq \underline{d}$, which implies that all $f_k, k \leq k_2$, are nondictatorial.

Claim 1:

$$k_2 \leq \underline{d}.$$

If $\underline{d} = n$ this is trivial. Suppose that $\underline{d} < n$. Let $l := n - \underline{d}$. As f is anonymous $\underline{d} > \frac{n}{2} \geq l$. Consider the following three agent social choice function. Let $g : P^{\{1,2,3\}} \to A$ be given by $g(p) := f(p(1)^{\underline{d}-l}, p(2)^l, p(3)^l)$, where $(p(1)^{\underline{d}-l}, p(2)^l, p(3)^l)$ can be understood similarly to $(p(1)^k, p(2)^l)$ above. As f is anonymous g is well defined and as f is unanimous, g is unanimous. Furthermore, as $f_{\underline{d}}$ is dictatorial, g is $\{1,2\}$ and $\{1,3\}$−dictatorial. So we can apply Lemma 70 to g.

If part (a) of Lemma 70 holds, i.e. if agent 1 is a dictator, then $f_{\tilde{d}}$, where $\tilde{d} := \max\{2l, \underline{d} - l\} \in K$, is dictatorial. If $\tilde{d} < \underline{d}$ then, as $\tilde{d} \geq \frac{n}{2}$, this contradicts the minimality of $\underline{d}$. So, $\tilde{d} \geq \underline{d}$, which implies that $2l \geq \underline{d}$. If $2l = \underline{d}$ then g is anonymous, contradicting dictatoriality. So, $l > \underline{d} - l$. Consider the profiles

$$r^i := (abc^{\underline{d}-l+i}, bca^{l-i}, cab^l), i \in \{0, \ldots, 2l - \underline{d}\}.$$

Then, $f(r^0) = a$, $f(r^{2l-\underline{d}}) = b$, as f is anonymous and g is dictatorial with dictator 1. But any r^{i+1} is a j-deviation of r^i for an agent j such that $(b, a) \in r^i(j)$. Hence, one of the r^i is manipulable. Clearly, $r^i \in M_f^3$ and $k(r^i) \leq \underline{d}$ for this r^i. For the same reasons a $\bar{r}^j$,

$$\bar{r}^j := (acb^{\underline{d}-l+j}, bca^{l-j}, cab^l), j \in \{0, \ldots, 2l - \underline{d}\},$$

is manipulable, in M_f^3, and $k(\bar{r}^j) \leq \underline{d}$ for this $\bar{r}^j$. As this $\bar{r}^j$ is anonymously inequivalent to r^i we have $k_2 \leq \max\{k(r^i), k(\bar{r}^j)\} \leq \underline{d}$.

So, suppose that part (b) of Lemma 70 is true. Then we have anonymously inequivalent $p, q \in M_g$, manipulable towards $(p_{-i}, t_p), (q_{-j}, t_q), i, j \in \{1, 2, 3\}$, respectively, such that $|p(N)| = |q(N)| = 3$ and $p(N) \cup \{t_p\} \neq q(N) \cup \{t_q\}$. If $i \in \{2, 3\}$ then by anonymity of f $i = 2$ without loss of generality. The same holds for j. Let

$$r^s := \begin{cases} (p(1)^{\underline{d}-l-s}, t_p^s, p(2)^l, p(3)^l) & \text{if } i = 1, s \in \{0, \ldots, \underline{d} - l - 1\}, \\ (p(1)^{\underline{d}-l}, p(2)^{l-s}, t_p^s, p(3)^l) & \text{if } i = 2, s \in \{0, \ldots, l - 1\}, \end{cases}$$

and

$$\bar{r}^u := \begin{cases} (q(1)^{\underline{d}-l-s}, t_q^s, q(2)^l, q(3)^l) & \text{if } j = 1, u \in \{0, \ldots, \underline{d} - l - 1\}, \\ (q(1)^{\underline{d}-l}, q(2)^{l-s}, t_q^s, q(3)^l) & \text{if } j = 2, u \in \{0, \ldots, l - 1\}. \end{cases}$$

By manipulability of p and q under g there are such r^s and $\bar{r}^u$ that are manipulable under f. As $|p(N)| = |q(N)| = 3$, $r^s, \bar{r}^u \in M_f^3$. As p and q are anonymously inequivalent and $p(N) \cup \{t_p\} \neq q(N) \cup \{t_q\}$ we have $[r^s] \nsim [\bar{r}^u]$. So, $k_2 \leq \max\{k(r^s), k(\bar{r}^u)\} \leq \underline{d}$. This finishes the proof of claim 1.

Let $k \in \{\lceil \frac{n}{2} \rceil, \ldots, k_2 - 1\}$. By claim 1 f_k is nondictatorial. So, by Theorem 68 there are $p, q \in M_{f_k}$. Note that we do not demand $p \neq q$ here, because we want to allow for the moment for the possibility that p and q fall under case (a) of Lemma 69. However, if necessary we can always choose p and q to satisfy $p \neq q$ by the same theorem. By the first part of Lemma 69 there are $(i, t_p) \in D_{f_k}(p), (j, t_q) \in D_{f_k}(q)$ such that $t_p \notin p(N), t_q \notin q(N)$. Let

$$r^s := \begin{cases} (p(1)^{k-s}, p(2)^l, t_p^s) & \text{if } i = 1, s \in \{0, \ldots, k-1\}, \\ (p(1)^k, p(2)^{l-s}, t_p^s) & \text{if } i = 2, s \in \{0, \ldots, l-1\}, \end{cases}$$

and

$$\bar{r}^u := \begin{cases} (q(1)^{k-u}, q(2)^l, t_q^u) & \text{if } i = 1, u \in \{0, \ldots, k-1\}, \\ (q(1)^k, q(2)^{l-u}, t_q^u) & \text{if } i = 2, u \in \{0, \ldots, l-1\}. \end{cases}$$

By manipulability of p and q under f_k there are s and u such that $r^s, \bar{r}^u \in M_f$ that are manipulable towards r^{s+1} or $\bar{r}^{u+1}$ respectively. Let $\bar{s}$ be the smallest index s such that an r^s is manipulable, i.e. the manipulation does not have to occur towards r^{s+1}. Let $\bar{u}$ be the smallest index u such that $\bar{r}^u$ is manipulable towards $\bar{r}^{u+1}$. The reason for these choices will become apparent in case (c).

Suppose that

$$k \in \{\left\lceil \frac{n}{2} \right\rceil, \ldots, k_1 - 1\}.$$

As $k(r^s) \leq k < k_1$, $k(\bar{r}^u) \leq k < k_1$, $|r^s(N)| \geq 3$ if $s \geq 1$ and $|\bar{r}^u(N)| \geq 3$ if $u \geq 1$, we must have $\bar{s} = 0$ and $\bar{u} = 0$. So $q_1^k := (p(1)^k, p(2)^l) = r^0 \in M_f$ and $q_2^k := (q(1)^k, q(2)^l) = \bar{r}^0 \in M_f$ satisfy $q_1^k, q_2^k \in M_f$, $|q_1^k(N)| = |q_2^k(N)| = 2$ and $k(q_1^k) = k(q_2^k) = k$. If $k = \frac{n}{2}$ we are done. If $k \neq \frac{n}{2}$ then $k > l$ and choosing $p \neq q$ implies that q_1^k and q_2^k are anonymously inequivalent. So, for $k \in \{\lceil \frac{n}{2} \rceil, \ldots, k_1 - 1\}$ we are done.

Suppose that

$$k \in \{k_1, \ldots, k_2 - 1\}.$$

We are done if $\bar{s} = 0$ or $\bar{u} = 0$. So, suppose contrapositive that $\bar{s} > 0$ and $\bar{u} > 0$. Using cases (a) to (c) of Lemma 69 we show that p and q can be chosen in such a way that $r^{\bar{s}} \sim \bar{r}^{\bar{u}}$. As $k(r^{\bar{s}}) \leq k$, $k(\bar{r}^{\bar{u}}) \leq k$, $|r^{\bar{s}}(N)| \geq 3$ and $|\bar{r}^{\bar{u}}(N)| \geq 3$ if $u \geq 1$, this contradicts $k < k_2$.

Case (a): We can choose p and q in M_{f_k} that are manipulable by different agents, say without loss of generality p by agent 1 and q by agent 2. Note that here we have to allow for $p = q$. Then

$$\max_{t \in P} |(r^s)^{-1}(t)| = \max\{k - \bar{s}, l, \bar{s}\} \leq k - 1 < k = (\bar{r}^{\bar{u}})^{-1}(q(1)).$$

so $r^{\bar{s}} \nsim \bar{r}^{\bar{u}}$.

Case (b): We can choose p and q in M_{f_k} such that $p(N) \cup \{t_p\} \neq q(N) \cup \{t_q\}$. Then $r^{\bar{s}} \nsim \bar{r}^{\bar{u}}$ is obvious.

Case (c): Either $i = j = 1$ and

$$p = (bca, cab), (1, cba) \in D_{f_k}(p)$$
$$q = (cba, bca), (1, cab) \in D_{f_k}(q)$$
$$f_k(p) = a$$

or $i = j = 2$ and

$$p = (cab, bca), (2, cba) \in D_{f_k}(p)$$
$$q = (bca, cba), (2, cab) \in D_{f_k}(q)$$
$$f_k(p) = a.$$

Suppose that $i = j = 1$ and $r^{\bar{s}} \sim \bar{r}^{\bar{u}}$. Then

$$r^{\bar{s}} = (bca^{k-\bar{s}}, cba^{\bar{s}}, cab^{l}) \sim (cba^{k-\bar{u}}, cab^{\bar{u}}, bca^{l}) = \bar{r}^{\bar{u}}.$$

This implies that $k - \bar{s} = l$, $\bar{s} = k - \bar{u}$ and $l = \bar{u}$. Note that $f(r^0) = f_k(p) = a$ is the least preferred alternative of the manipulating agent, who has the preference bca. By the minimality of $\bar{s}$ and the anonymity of f this implies that $f(r^0) = f(r^1) = \ldots f(r^{\bar{s}-1}) = f(r^{\bar{s}}) = f(\bar{r}^{\bar{u}}) = a$. Furthermore, as $(cba^{k-\bar{u}}, cab^{\bar{u}}, bca^{l})$ is manipulable by an agent with preference cba, we have $f(cba^{k-\bar{u}-1}, cab^{\bar{u}+1}, bca^{l}) \in \{b, c\}$. Hence,

$$f(cba^{k-\bar{u}-1}, cab^{\bar{u}+1}, bca^{l}) = f(cba^{\bar{s}-1}, cab^{l+1}, bca^{k-\bar{s}}) \in \{b, c\}.$$

But then $r^{\bar{s}-1}$ is manipulable, as $f(r^{\bar{s}-1}) = f(bca^{k-\bar{s}+1}, cba^{\bar{s}-1}, cab^{l}) = a$, and by deviating to cab any agent with preference bca can manipulate. This contradicts the minimality of $\bar{s}$.

Suppose that $i = j = 2$ and $r^{\bar{s}} \sim \bar{r}^{\bar{u}}$. Then

$$r^{\bar{s}} = (bca^{k}, cab^{l-\bar{s}}, cba^{\bar{s}}) \sim (cba^{k}, bca^{l-\bar{u}}, cab^{\bar{u}}) = \bar{r}^{\bar{u}}.$$

This implies by $k > l$ that $bca = cba$, a contradiction.

Hence, in any of the cases (a) to (c) p and q can be chosen in such a way that $r^{\bar{s}} \nsim \bar{r}^{\bar{u}}$. This contradicts then $k < k_2$ if $\bar{s} > 0$ and $\bar{u} > 0$. So, without loss of generality $\bar{u} = 0$, and letting $q_1^k = r^0 \in M_f$ we have $|q_1^k(N)| = 2$ and $k(q_1^k) = k$.

This finishes the proof of the lemma. ∎

By the remarks before Lemma 72 we can now state the following theorem.

Theorem 73 Let $f : P^N \to A$ be a unanimous and anonymous social choice function. Then $|M_f| \geq 2^n - 2$.

7.5 Social choice functions f^* that attain the lower bound $2^n - 2$

Let $F^* := \{f \in F \mid |M_f| = 2^n - 2\}$ be the set of all social choice functions attaining the lower bound. In this section we formulate a conjecture about F^*, saying that $f^* \in F^*, n \geq 4$, if and only if f^* belongs to one of three classes of social choice functions which we describe completely. There will be an exception for $n \in \{2, 3\}$. First of all one class is only defined for $n \geq 3$, where its definition in the case $n = 3$ compared to $n \geq 4$ alters a litte. Furthermore, for $n = 3$ a fourth extra class of social choice functions is needed. Altogether we define four classes $C_i, i \in \{1, 2, 3, 4\}$, of social choice functions such that $C_i \subseteq F^*$. For any $x \in A$ and $p \in P^N$ let

$$Pa(x, p) := \{y \in A - \{x\} \mid (y, x) \in p(i) \text{ for all } i \in N\}$$

be the set of all alternatives that *Pareto dominate* x in p. In 1. up to 5. it is to be understood that indefiniteness is resolved in such a way that anonymity and unanimity hold. Let $A = \{x, y, z\}$.

1. Let $C_1 \subseteq F$ be the set of all social choice functions f such that

$$f(p) = \begin{cases} z & \text{if } Pa(z, p) = \varnothing \\ x & \text{if } Pa(z, p) = \{x\} \\ y & \text{if } Pa(z, p) = \{y\} \end{cases}$$

   and $f(p) = f(q) \in \{x, y, z\}$ for all $q \in P^N$ such that

$$Pa(z, p) = Pa(z, q) = \{x, y\}, [q] = [p].$$

2. Let $C_2 \subseteq F$ be the set of all social choice functions f such that

$$f(p) = \begin{cases} z & \text{if } Pa(z, p) = \varnothing \\ x & \text{if } Pa(z, p) = \{x\} \\ y & \text{if } Pa(z, p) = \{y\} \text{ and } p \notin [(xyz^{n-1}, yzx)] \\ z & \text{if } p \in [(xyz^{n-1}, yzx)] \\ x & \text{if } p \in [(xyz^{n-1}, yxz)] \end{cases}$$

   and $f(p) = f(q) \in \{x, y, z\}$ for all $p, q \in P^N$ such that $Pa(z, p) = Pa(z, q) = \{x, y\}, [p] = [q] \neq [(xyz^{n-1}, yxz)]$.

3. If $n \geq 4$ let $C_3 \subseteq F$ be the set of all social choice functions f such that

$$f(p) = \begin{cases} z & \text{if } Pa(z, p) = \varnothing \\ x & \text{if } Pa(z, p) = \{x\} \text{ and } p \notin [(yxz^{n-1}, xzy)] \\ y & \text{if } Pa(z, p) = \{y\} \text{ and } p \notin [(xyz^{n-1}, yzx)] \\ z & \text{if } p \in [(yxz^{n-1}, xzy)] \\ z & \text{if } p \in [(xyz^{n-1}, yzx)] \\ x & \text{if } p \in [(xyz^{n-1}, yxz)] \cup [(xyz^{n-2}, yxz^2)] \\ y & \text{if } p \in [(yxz^{n-1}, xyz)] \cup [(yxz^{n-2}, xyz^2)] \end{cases}$$

and $f([p]) \in \{x, y, z\}$ if $Pa(z, p) = \{x, y\}$ and $p \notin [(xyz^{n-1}, yxz)] \cup [(xyz^{n-2}, yxz^2)] \cup [(yxz^{n-1}, xyz)] \cup [(yxz^{n-2}, xyz^2)]$.

4. If $n = 3$ let $C_3 \subseteq F$ be the set of all social choice functions f such that

$$
f(p) = \begin{cases}
z & \text{if } Pa(z, p) = \varnothing \\
x & \text{if } Pa(z, p) = \{x\} \text{ and } p \notin [(yxz^2, xzy)] \\
y & \text{if } Pa(z, p) = \{y\} \text{ and } p \notin [(xyz^2, yzx)] \\
z & \text{if } p \in [(yxz^2, xzy)] \\
z & \text{if } p \in [(xyz^2, yzx)] \\
x & \text{if } p \in [(xyz^2, yxz)] \\
y & \text{if } p \in [(yxz^2, xyz)]
\end{cases}
$$

and $f(p) \in \{x, y, z\}$ if $Pa(z, p) = \{x, y\}$ and $p \notin [(xyz^2, yxz)] \cup [(yxz^2, xyz)]$.

5. If $n = 3$ let $C_4 \subseteq F$ be the set of all social choice functions f for which there are $x, y \in A, x \neq y$, such that $f(p)$ is determined by majority voting between x and y, i.e.

$$
f(p) = \begin{cases}
x & \text{if } |\{i \in N \mid (x, y) \in p(i)\}| > |\{i \in N \mid (y, x) \in p(i)\}|, \\
y & \text{if } |\{i \in N \mid (y, x) \in p(i)\}| > |\{i \in N \mid (x, y) \in p(i)\}|,
\end{cases}
$$

for all profiles p where $f(p)$ is not determined already by unanimity.

The following lemma can be proven by counting manipulable profiles and evaluating the freedom left in the definitions of the C_i. The counting argument can be simplified by noting that no agent will want to manipulate to a Pareto dominated alternative. Since the rest is elementary counting we do not present the proof here.

Lemma 74 *Let F^* be the set of all minimally manipulable unanimous and anonymous social choice function with n agents and three alternatives. Then, $C_1 \cup C_2 \subseteq F^*$ if $n = 2$, $C_1 \cup C_2 \cup C_3 \cup C_4 \subseteq F^*$ if $n = 3$ and $C_1 \cup C_2 \cup C_3 \subseteq F^*$ if $n \geq 4$. Furthermore, the C_i that are defined in a particular case are disjoint. We have $|C_1| = 3^n$, $|C_2| = 2 * 3^{n-1}$, and $|C_3| = 3^{n-2}$, $|C_4| = 3$, if they are defined, and*

$$
|F^*| \geq \begin{cases}
15 & \text{if } n = 2, \\
51 & \text{if } n = 3, \\
3^n + 2 * 3^{n-1} + 3^{n-2} & \text{if } n \geq 4.
\end{cases}
$$

We conjecture that equalities hold in the lemma. We believe that the proof of the main result in Section 7.4 can be used to show this conjecture and explain how this might be done.

First of all we note that in the statement of Lemma 71

$$
[p_i] \geq \sum_{j=0}^{l(p_i)-1} \binom{n}{k(p_i) + j}, p_i \in M_f^3,
$$

equality cannot hold. Going through the proof of that statement one can see that equality would imply that $l(p_i) = 1$, $k(p_i) = n - 1$, and $|p_i(N)| = 2$, a contradiction. So, $M^3_{f*} = \emptyset$ for any $f^* \in F^*$. This also implies that $\underline{d} = n$. Hence, all manipulable profiles arise from the f_k. In particular for each f_k we must have that it yields only two anonymously inequivalent manipulable profiles. Now, classifying all such f_k, we know the possible M_f's and what f can look like on $\{p \in P^N \mid |p(N)| \leq 2\}$. Connecting this information by strategy-proofness, in particular on $B^3 := \{p \in P^N \mid |p(N)| \geq 3\}$, it might be possible to characterize F^*. The reason why we do not pursue this here is the classification of all such f_k. There are 125 unanimous and nondictatorial two agent social choice functions with two manipulable profiles which have to be checked first, and there might be even more with more manipulable profiles, but only two that give anonymously inequivalent profiles for f.

We mention also that we have confirmed the following conjecture up to 8 agents by computer, which is also how we found the 125 unanimous and nondictatorial two agent social choice functions over three alternatives.

Conjecture 75 *Let F^* be the set of all minimally manipulable unanimous and anonymous social choice functions for a fixed size of the agent set n. Then*

$$F^* = \begin{cases} C_1 \cup C_2 & \text{if } n = 2, \\ C_1 \cup C_2 \cup C_3 \cup C_4 & \text{if } n = 3, \\ C_1 \cup C_2 \cup C_3 & \text{if } n \geq 4, \end{cases}$$

where the C_i are defined as in Lemma 74. Furthermore, as the C_i are disjoint,

$$|F^*| = \begin{cases} 9 + 6 = 15 & \text{if } n = 2, \\ 27 + 18 + 3 + 3 = 51 & \text{if } n = 3, \\ 3^n + 2 * 3^{n-1} + 3^{n-2} & \text{if } n \geq 4. \end{cases}$$

7.6 Pareto optimality and anonymity

In this section we show how our results and the conjecture apply when unanimity is replaced by the stronger requirement of Pareto optimality.

A social choice function is called *Pareto optimal* if it does not choose Pareto dominated alternatives, i.e. if $Pa(x, p) \neq \emptyset$, $x \in A$, implies that $f(p) \neq x$. Pareto optimality is a stronger requirement than unanimity. Let p be such that $f(p) = x$ is determined by unanimity. Then $Pa(y, p) \supseteq \{x\}$ for all $y \in A - \{x\}$, implying that $f(p) \neq y$ for all $y \in A - \{x\}$. So $f(p) = x$ also by Pareto optimality. On the other hand $f(abc, bca) \in \{a, b\}$ by Pareto optimality, but unanimity does not reduce the choice at this profile. Let G be the set of Pareto optimal and anonymous social choice functions. Then $G \subset F$.

Let $\overline{C_1} \subseteq G$ be the set of all social choice functions $g \in G$ for which there is a labeling

of the alternatives $A = \{x, y, z\}$ such that

$$g(p) = \begin{cases} z & \text{if } Pa(z, p) = \emptyset \\ x & \text{if } Pa(z, p) = \{x\} \\ y & \text{if } Pa(z, p) = \{y\} \end{cases}$$

and $g(p) = g(q) \in \{x, y\}$ for all $q \in P^N$ such that

$$Pa(z, p) = Pa(z, q) = \{x, y\}, [q] = [p].$$

Note that $\overline{C_1} \subset G \subset F$ and $\overline{C_1} \subset C_1 \subset F^*$ So the following lemma is implied by Lemma 74 and counting $\overline{C_1}$.

Lemma 76 *Let G^* be the set of minimally manipulable social choice functions in G. Then*

$$\overline{C_1} \subseteq G^*, \; |\overline{C_1}| = 3 * 2^n \text{ and } |G^*| \geq 3 * 2^n.$$

Furthermore, $G \subseteq F$ implies then the following corollary to Theorem 73.

Corollary 77 *Let $g : P^N \to A$ be a Pareto optimal and anonymous social choice function. Then $|M_g| \geq 2^n - 2$.*

It is easily checked that there are no Pareto optimal social choice functions in $C_i, i \in \{2, 3, 4\}$, whenever these sets are defined. So we can state the following conjecture based on conjecture 75.

Conjecture 78 *Let G^* be the set of all minimally manipulable Pareto optimal and anonymous social choice functions for a fixed size of the agent set n. Then*

$$G^* = \overline{C_1},$$

and

$$|G^*| = 3 * 2^n.$$

Note that for any $g \in \overline{C_1}$ there is an alternative z that is chosen whenever it is not Pareto dominated by one of the other alternatives. We call this alternative the *status quo*. If z is Pareto dominated, then the choice is taken among one of the dominating alternatives in an arbitrary way. In practice this arbitrariness might be replaced by a voting rule between x and y that seems reasonable in the given situation. For example it might be that y is declared as a second status quo, so that x is only chosen if $p = x \dots$".

Anyway these explanations show that our result has a nice interpretation if a status quo z among the alternatives exists, e.g. a current jurisdiction z about which there is consensus that it should only be altered to one of two new jurisdictions x and y if every voter prefers the new jurisdiction. Then, the least manipulable social choice functions that guarantee anonymity and Pareto optimality of the outcome are as follows. Choose z in the situation where there is consensus that no change should be made. In the other situations choose an alternative Pareto dominating z; if necessary, i.e. if there are two such alternatives, according to any unanimous social choice function over two alternatives.

References

Akerlof, G.A. (1970): "The market for "Lemons": Quality uncertainty and the market mechanism". Quarterly Journal of Economics 84, 488-500.

Aleskerov, F., and E. Kurbanov (1999): "Degree of manipulability of social choice procedures ". In: Proceedings of the Third International Meeting of the Society for the Advancement of Economic Theory. Berlin/Heidelberg/New York: Springer.

Allen, B. (1984): "Generic existence of completely revealing equilibria for economies with uncertainty when prices convey information". Econometrica 49, 1173-1199.

Allen, B. (1991a): "Market games with asymmetric information and non-transferable utility: Representation results and the core". Center for Analytic Research in Economics and the Social Sciences, Working Paper #91-09, University of Pennsylvannia.

Allen, B. (1991b): "Transferable utility market games with asymmetric information: Representation results and the core". Center for Analytic Research in Economics and the Social Sciences, Working Paper #91-11, University of Pennsylvannia.

Allen, B. (1994): "Market games with asymmetric information: The core with finitely many states of the world". In B. Munier and M. J. Machina (Eds.), Models and Experiments in Risk and Rationality. Dordrecht: Kluwer Academic Press.

Arrow, K. (1963): "Social choice and individual values". New York: Wiley.

Dasgupta, P., Hammond, P., and E. Maskin (1979): "The implementation of social choice rules: Some general results on incentive compatibility". Review of Economic Studies 46, 185-216.

Debreu, G. (1959): "Theory of Value". Cowles Foundation Monograph 17. New York: Wiley.

Einy, E., Moreno, D., and B. Shitovitz (2000): "On the core of an economy with differential information". Journal of Economic Theory 94, 262-270.

Einy, E., Moreno, D., and B. Shitovitz (2001): "Competitive and core allocations in large economies with differential information". Economic Theory 18, 321-332.

Fristrup, P., and H. Keiding (1998): "Minimal manipulability and interjacency for two-person social choice functions". Social Choice and Welfare 15, 455-467.

Gibbard, A. (1973): "Manipulation of voting schemes: a general result". Econometrica 41, 587-602.

Harsanyi, J.C. (1959): "A Bargaining Model for the Cooperative n-person Game". In: Tucker, A.W. and Luce, R.D., (Eds.), Contributions to the Theory of Games IV, Annals of Mathematics Studies No. 40. Princeton/New Jersey: Princeton University Press.

Hart, S., and A. Mas-Colell (1989): "Potential, Value, and Consistency". Econometrica 57, 589-614.

Hervés-Beloso, C., Moreno-García, E., and N. Yannelis (2003): "Characterizing Radner Equilibria". Working paper.

Hildenbrand, W., and A.P. Kirman (1988): "Equilibrium Analysis. Variations on themes by Edgeworth and Walras". Advanced Textbooks in Economics, Vol. 28. Amsterdam/New York/Oxford: North Holland.

Housman, D. (2001): "Linear and symmetric allocation methods for partially defined co-operative games". International Journal of Game Theory 30, 377-404.

Kelly, J.S. (1988): "Minimal manipulability and local strategy-proofness". Social Choice and Welfare 5, 81-85.

Kelly, J.S. (1989): "Interjacency". Social Choice and Welfare 6, 331-355.

Kelly, J.S. (1993): "Almost all social choice rules are highly manipulable, but a few aren't". Social Choice and Welfare 10, 161-175.

Koutsougeras, L., and N. Yannelis (1993): "Incentive compatiblity and information superiority of the core of an economy with differential information". Economic Theory 3, 195-216.

Krasa, S., and N. Yannelis (1994): "The value allocation of an economy with differential information". Econometrica 62, 881-900.

Krasa, S., and N. Yannelis: "Existence and properties of a value allocation for an economy with differential information". Journal of Mathematical Economics 25, 165-179.

Liu, C. (1992): "From a theorem of Radner". Economics Letters 39, 401-404.

Maus, S. (2002): "Balancedness and the core in economies with asymmetric information". Economic Theory 22, 613-627.

Maus, S. (2003): "The expected Shapley value". Research Memoranda Universiteit Maastricht RM /03/004, University of Maastricht.

Maus, S. (2004a): "Exchange economies with asymmetric information: competitive equilibrium and core". Economic Theory 24, 395-418.

Maus, S., Peters, H., and A. Storcken (2004b): "Minimal manipulability: Anonymity and Surjectivity". Research Memoranda Universiteit Maastricht RM /04/007, University of Maastricht.

Maus, S., Peters, H., and A. Storcken (2004c): "Minimal manipulability: Unanimity and Nondictatorship". Research Memoranda Universiteit Maastricht RM /04/006, University of Maastricht.

Maus, S., Peters, H., and A. Storcken (2004d): "Minimal manipulability: Anonymity and Unanimity". Research Memoranda Universiteit Maastricht RM /04/0026, University of Maastricht.

May, K.O. (1952): "A Set of Independent, Necessary and Sufficient Conditions for Simple Majority voting". Econometrica 20, 680-684.

Moulin, H. (1983): "The strategy of social choice". Advanced Textbooks in Economics 18. Amsterdam/New York/Oxford: North-Holland.

Page, F. (1997): "Market games with differential information and infinite dimensional commodity spaces: The core". Economic Theory 9, 151-159.

Radner, R. (1968): "Competitive Equilibrium under Uncertainty". Econometrica 36, 31-58.

Radner, R. (1979): "Rational expectations equilibrium: Generic existence and the information revealed by prices". Econometrica 40, 655-678.

Radner, R. (1982): "Equilibrium under Uncertainty". In: Arrow, K.J. and Intrilligator, M.D. (Eds.), Handbook of Mathematical Economics Vol. II. Amsterdam/New York/Oxford: North-Holland.

Satterthwaite, M. (1975): "Strategy-proofness and Arrow's conditions: existence and correspondence theorems for voting procedures and social welfare functions". Journal of Economic Theory 10, 187-217.

Scarf, H.E. (1967): "The core of an n-person game". Econometrica 35, 50-69.

Schwalbe, U. (1999): "The core of economies with asymmetric information". In: Lecture Notes in economics and mathematical systems 474. Berlin/Heidelberg/New York: Springer.

Shapley, L.S. (1953): "A Value for n-person games". In: Kuhn, H.W. and Tucker, A.W., (Eds.), Contributions to the Theory of Games II, Annals of Mathematics Studies No. 28. Princeton/New Jersey: Princeton University Press.

Slinko, A. (2002): "On asymptotic strategy-proofness of classical social choice rules". Theory and Decision 52, 389-398.

Willson, S.J. (1993): "A Value for Partially Defined Cooperative Games". International Journal of Game Theory 21, 371-384.

Wilson, R. (1978): "Information, Efficiency and the Core of an Economy". Econometrica 46, 807-816.

Yannelis, N. (1991): "The Core of an Economy with Differential Information". Economic Theory 1, 183-198.

Young, H.P. (1985): "Monotonic Solutions of Cooperative Games". International Journal of Game Theory 14, 65-72.

Author Index

Shapley, L.S. (1953). "A Value for n-person games". In: Kuhn, H.W. and ... W.
(Eds.), Contributions to the Theory of Games II, Annals of Mathematics ... 28.
Princeton/New ... Princeton University Press.

Sinico, A (2003). ... strategy-proofness of classical social choice ...
ory and Decision ...

Wilson, ... Defined Cooperative Game ...
Journal ...

Wilson, R. (19 ... and the Core of an Economy ...
46, 807-816

Yannelis, N ... conomy with Differential Information ...
Theory 1, 183 ...

Young, H.P. ... Cooperative Games ...
of Game Te ... 78-80, 90, 91, 97,

About the Author

Stefan Maus was born on December 22, 1975 in Erkelenz, Germany. From April 1996 till January 2001 he studied mathematics at the RWTH Aachen, Germany. In between he spent a year at the Imperial College, London, UK, where he received the Imperial College International Diploma, under the Erasmus exchange program from October 1997 to September 1998. In January 2001 he received the master's degree in mathematics (German title: "Diplom-Mathematiker (Univ.))". Professor Dr. Hans Peters supervised his master's thesis.

From February 2001 on he worked as an "assistent in opleiding" (AIO) at the Department of Quantitative Economics at Maastricht University, The Netherlands. His research in economic theory, game theory and social choice theory was supervised by Professor Dr. Hans Peters and Dr. Ton Storcken.

bc,acb,: b;c;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;
bc,bac,: b;b;c;c;c;c;c;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b
bc,bca,: b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b
bc,cab,: b;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;b;b;b;b;b;b;b;b;a;a
bc,cba,: b;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;b;b;b;b;b;b;b;b;b
cb,abc,: b;c;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a
cb,acb,: a;c;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a;a
cb,bac,: b;c;a;a;a;a;a;c;c;c;c;c;b;b;b;b;b;b;b;b;b;b;b;a;a;a;a;a;a;a;a;a;c;b;b;b;b;b;b;a;c
cb,bca,: b;c;c;c;c;c;c;c;c;c;c;c;b;b;b;b;b;b;b;b;b;b;b;c;c;c;c;c;c;c;c;c;b;b;b;b;b;b;c;c
cb,cab,: c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;a
cb,cba,: c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c
bac,abc,: b;b;c;b;b;a;a;c;b;b;a;a;b;b;b;b;b;b;b;a;a;a;c;b;b;a;a;a;a;a;b;b;b;b;b;b;b;b
bac,acb,: b;c;a;c;a;a;a;a;c;a;a;a;a;c;c;c;c;a;a;a;a;a;a;a;c;a;a;a;a;a;a;b;b;b;b;b;b;b
bac,bac,: b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b
bac,bca,: b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b
bac,cab,: b;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;b;b;b;b;b;b;b;b
bac,cba,: b;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;b;b;b;b;b;b;b;b
cba,abc,: b;b;b;b;b;c;b;b;b;b;b;c;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;c;b;a;a;a;b;b;b;b;b;b;b
cba,acb,: b;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;a;a;a;b;b;b;b;b;b;b
cba,bac,: b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b
cba,bca,: b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b;b
cba,cab,: b;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;b;b;b;b;b;b;b
cba,cba,: b;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;b;b;b;b;b;b;b
ab,abc,: b;c;c;c;c;c;c;c;c;c;c;c;b;b;a;a;b;b;a;a;b;a;a;a;c;c;c;c;c;a;a;a;c;b;a;a;a;a;c;c;c
ab,acb,: c;c;c;c;c;c;c;c;c;c;c;c;c;c;a;c;c;a;c;c;a;c;c;c;c;a;a;a;c;c;c;b;a;a;c;c;c
ab,bac,: b;c;c;c;c;c;c;c;c;c;c;c;b;b;b;b;b;b;b;b;b;c;c;c;c;c;a;a;a;c;b;b;b;b;b;c;c
ab,bca,: b;c;c;c;c;c;c;c;c;c;c;c;b;b;b;b;b;b;b;b;b;c;c;c;c;c;c;c;c;b;b;b;b;b;b;c;c
ab,cab,: c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c
ab,cba,: c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c
ba,abc,: b;c;c;c;c;c;c;c;c;c;c;c;c;b;b;b;b;b;b;b;b;b;b;c;c;c;c;c;a;a;a;c;b;b;b;b;b;c;c
ba,acb,: c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;a;a;a;c;c;c;c;b;c;c;c
ba,bac,: b;c;c;c;c;c;c;c;c;c;c;c;c;b;b;b;b;b;b;b;b;b;c;c;c;c;c;b;a;c;b;b;b;b;b
ba,bca,: b;c;c;c;c;c;c;c;c;c;c;c;c;b;b;b;b;b;b;b;b;b;c;c;c;c;c;b;c;c;b;b;b;b;b
ba,cab,: c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c;c