Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI 2015)

Equilibrium Refinement through Negotiation in Binary Voting

Umberto Grandi
University of Toulouse
France
umberto.grandi@irit.fr

Davide Grossi
University of Liverpool
United Kingdom
d.grossi@liverpool.ac.uk

Paolo Turrini
Imperial College London
United Kingdom
paolo.turrini@imperial.ac.uk

Abstract

We study voting games on binary issues, where voters might hold an objective over some issues at stake, while willing to strike deals on the remaining ones, and can influence one another's voting decision before the vote takes place. We analyse voters' rational behaviour in the resulting two-phase game, showing under what conditions undesirable equilibria can be removed as an effect of the prevote phase.

1 Introduction

Social choice theory, and voting theory in particular, have been gaining increased attention in the multiagent systems (MAS) literature in the last decade, and voting is considered a fundamental tool for the study of MAS [Brandt *et al.*, 2014].

In the face of the extreme popularity of the voting paradigm, the MAS literature studying voting as a fully-fledged form of strategic interaction, i.e., as a non-cooperative game, is very small (although growing, e.g., [Desmedt and Elkind, 2010; Xia and Conitzer, 2010; Obraztsova *et al.*, 2013]). In particular, no work with the notable exception of the literature on iterative voting [Meir *et al.*, 2010; Lev and Rosenschein, 2012; Brânzei *et al.*, 2013] has studied how voting behavior in rational agents is influenced by strategic forms of interaction that precede the voting stage. Literature in social choice has recognised that interaction preceding voting can be an effective tool to induce opinion change and achieve compromise solutions [Dryzek and List, 2003; List, 2011] while in game theory pre-play negotiations are known to be effective in overcoming inefficient allocations caused by players' individual rationality [Jackson and Wilkie, 2005]. When players are allowed to offer a part of their gains at certain outcomes to influence the decisions of the other agents, they are able to overcome highly inefficient scenarios, such as the Prisoners' Dilemma [Jackson and Wilkie, 2005].

In this paper we study pre-vote negotiations in voting games over binary (yes/no) issues, where voters hold a special type of lexicographic preferences over the set of issues at stake, i.e., hold an objective about a subset of them while they are willing to negotiate on the remaining ones, and can influence one another before casting their ballots by transferring utility in order to obtain a more favourable outcome. We show that this type of pre-vote interaction has beneficial effects on voting games by refining their set of equilibria.

Related work Our approach relates directly to several ongoing lines of research in social choice, game theory and their applications to MAS.

Binary Aggregation and Voting Games. We study societies of voters that express a yes/no opinion on issues at stake. The setting is also known as *voting in multiple referenda* and closely related to the growing literature on voting games. Classical references include the work of Dhillon and Lockwood [2004] and Messner and Polborn [2007], and more recently lead to computational studies of best-response dynamics in voting games [Meir *et al.*, 2010; Xia and Conitzer, 2010; Lev and Rosenschein, 2012]. In binary voting, aside from the (non-)manipulability of voting rules (see, for instance, [Dietrich and List, 2007b]), non-cooperative game-theoretic aspects are underexplored and are our focus here. Binary voting can be further enriched by imposing that individual opinions also need to satisfy a set of integrity constraints, like in binary voting with constraints [Grandi and Endriss, 2013] and judgment aggregation [Dietrich and List, 2007a; Grossi and Pigozzi, 2014]. Standard preference aggregation, which is the classical framework for voting theory, is a special case of binary voting with constraints [Dietrich and List, 2007a]. Voting with constraints will be touched upon towards the end of the paper.

Boolean games. We model voting strategies in binary aggregation as boolean games [Harrenstein *et al.*, 2001; Wooldridge *et al.*, 2013], allowing voters to have control of a set of propositional variables, i.e., their ballot, and to assign utilities to outcomes, with specific goal outcomes they want to achieve. In our setting however goals of individuals are expressed on *the outcome* of the decision process, thus on outcomes that do not depend on their single choice only. Unlike boolean games, where each actor uniquely controls a propositional variable, in our setting the control of a variable is shared among the voters and its final truth value is determined by a voting rule.

Election control. The field of computational social choice has extensively studied lobbying [Christian *et al.*, 2007;

Bredereck *et al.*, 2014] and bribery [Baumeister *et al.*, 2013; Hazon *et al.*, 2013], modelled from the single agent perspective of a lobbyist or briber who tries to influence voters' decisions through monetary incentives, or from the perspective of a coalition of colluders [Bachrach *et al.*, 2011]. Here we study forms of control from a non-cooperative game-theoretic perspective where any voter can influence any other voter.

Equilibrium refinement. Non-cooperative models of voting are known to suffer from a multiplicity of equilibria, many of which appear counterintuitive. Equilibrium selection or refinement is a vast and long-standing research program in game theory [Meyerson, 1978]. Models of equilibrium refinement have been applied to voting games in the literature on economics [Gueth and Selten, 1991; Kim, 1996] and within MAS [Desmedt and Elkind, 2010; Obraztsova *et al.*, 2013], as well as the above mentioned iterative voting model, which offers a natural strategy for selecting equilibria through best response dynamics from the profile of truthful votes. In this paper we study a two-phase model for equilibrium refinement in a voting game where equilibria are selected by means of an initial pre-vote negotiation phase.

Pre-play negotiations. We model negotiations as a pre-play interaction phase, in the spirit of Jackson and Wilkie [2005]. During this phase, which precedes the play of a normal form game, players are entitled to sacrifice a part of their final utility in order to convince their opponents to play certain strategies, which in our case consist of voting ballots. In doing so we build upon and simplify the framework of endogenous boolean games [Turrini, 2013], which enriches boolean games with a pre-play phase.

Paper contribution and outline We describe a model of equilibrium refinement for voting games which: (i) is applicable to one-shot voting in the general context of binary aggregation; (ii) does not rely on limit behavior in repeated interactions; and (iii) can capture the compromise-seeking phase that typically precedes decision-making by voting. More specifically we address the effect of pre-play negotiations on the outcomes of voting games on binary (yes-no) issues. We isolate precise conditions under which bad equilibria – e.g., inefficient ones – can be overcome, and good ones sustained.

The paper is organised as follows. First, in Section 2 we present the setting of binary aggregation, defining the (issue-wise) majority rule and a more general class of aggregation procedures, which constitute the rules of choice for the current paper. Second, we define voting games for binary aggregation, specifying individual preferences by means of both a goal and a utility function, and we show how undesirable equilibria can be removed by appropriate modifications of the game matrix (Section 3). Third, we present a full-blown model of collective decisions as a two-phase game, with a negotiation phase preceding the vote. We show how the set of equilibria can be refined by means of rational negotiations removing undesirable equilibria and, dually, maintaining desirable ones (Section 4). Section 6 concludes.

2 Preliminaries

We model situations of collective decision-making in the framework of binary aggregation. In this setting a finite set of agents express yes/no opinions on a finite set of binary issues, and these opinions are then aggregated into a collective decision over each issue.

Definition 1 (*BA* structure). *A binary aggregation structure (BA structure) is a tuple $\mathcal{S} = \langle \mathcal{N}, \mathcal{I} \rangle$ where:*

- $\mathcal{N} = \{1, \dots, n\}$ *is a finite set individuals;*[1]
- $\mathcal{I} = \{1, \dots, m\}$ *is a finite set of issues.*

We denote $\mathcal{D} = \{B \mid B : \mathcal{I} \to \{0,1\}\}$ the set of all possible binary opinions over the set of issues $\mathcal{I}$ and call an element $B \in \mathcal{D}$ a *ballot*. Thus, $B(j) = 0$ (respectively, $B(j) = 1$) indicates that the agent who submits ballot B rejects (respectively, accepts) the issue j.

A *profile* $\boldsymbol{B} = (B_1, \dots, B_n)$ is the choice of a ballot for every individual in $\mathcal{N}$. We write B_i to denote the ballot of individual i within a profile $\boldsymbol{B}$. Thus, $B_i(j) = 1$ indicates that individual i accepts issue j in profile $\boldsymbol{B}$. Furthermore we denote by $\mathcal{N}_j^{\boldsymbol{B}} = \{i \in \mathcal{N} \mid B_i(j) = 1\}$ the set of individuals accepting issue j in profile $\boldsymbol{B}$.

Definition 2 (Aggregation rule). *Given a BA structure $\mathcal{S}$, an aggregation rule (or aggregator) for $\mathcal{S}$ is a function $F : \mathcal{D}^{\mathcal{N}} \to \mathcal{D}$, mapping every profile to a binary ballot in $\mathcal{D}$. $F(\boldsymbol{B})(j)$ denotes the outcome of the aggregation on issue j.*

Possibly the best-known aggregation rule is *issue-by-issue strict majority rule* (maj), which accepts an issue if and only if the majority of voters accept it, formally $maj(\boldsymbol{B})(j) = 1$ if and only if $|\mathcal{N}_j^{\boldsymbol{B}}| \geq \frac{|\mathcal{N}|+1}{2}$. Other notable examples of aggregation rules include *quota rules*, which accept an issue if the number of voters accepting it exceeds a possibly different quota for each issue, and *distance-based rules*, which output the ballot that minimises the overall distance to the profile for a suitable notion of distance.

Example 1. *A parliament composed by equally representative parties A, B, C is to decide whether to develop atomic weapons (W), importing nuclear technology from the foreign market (F), and build in-house nuclear plants (P). The profile in Table 1 is an instance of binary aggregation with the majority rule, in which each individual submits a binary opinion over each of the three issues at stake.*

	W	F	P
Party A	1	0	1
Party B	1	1	0
Party C	0	0	0
Majority	1	0	0

Table 1: An instance of binary aggregation

Given an aggregation rule F, we call a subset of voters $C \subseteq \mathcal{N}$ a *winning coalition* if for every profile $\boldsymbol{B}$, issues $j \in \mathcal{I}$ and $x \in \{0,1\}$, if $C = \{i \in \mathcal{N} \mid B_i(j) = x\}$ then $F(\boldsymbol{B})(j) = x$. We call C a *resilient winning coalition* if C is a winning coalition and $C \setminus \{i\}$ is also a winning coalition

[1]Throughout the paper we will make the implicit assumption that the set of individuals consists of at least three elements.

for every $i \in C$.[2] Given an aggregator F, we denote with $\mathcal{W}_F$ the set of winning coalitions for F, and with $\mathcal{W}_F^+$ the set of resilient winning coalitions. In the case of the majority rule we have $\mathcal{W}_{maj}^+ = \left\{C \subseteq \mathcal{N} \mid |C| \geq \frac{|\mathcal{N}|+1}{2} + 1\right\}$, i.e., all coalitions exceeding the majority threshold of at least one element are resilient.

Aggregation rules are classified by means of axioms that bind the properties of the outcome at certain profiles. An important example is the axiom of *unanimity*, which demands that the outcome of aggregation coincide with the individuals' judgments in case of consensus. We refer the reader to the relevant literature for a formal treatment of axiomatic properties. Here we provide just the following definitions in terms of winning coalitions:

Definition 3 (Systematicity). *An aggregator F is called* systematic *if it can be characterised through winning coalitions, i.e., if there exists a set $\mathcal{W}_F \subseteq 2^{\mathcal{N}}$ such that for all profiles $\boldsymbol{B}$ and issues $j \in \mathcal{I}$, we have that $F(\boldsymbol{B})(j) = 1$ iff $\mathcal{N}_j^{\boldsymbol{B}} \in \mathcal{W}_F$.*

Definition 4 (Monotonicity). *A systematic rule F is called* monotonic *if its set of winning coalitions is closed under supersets, i.e., for all $C \in \mathcal{W}_F$, if $C \subseteq C'$ then $C' \in \mathcal{W}_F$.*

The majority rule and all quota rules satisfy these axioms, but systematicity, for instance, is violated by most distance-based rules. In this paper we focus on systematic and monotonic rules, as a strict generalisation of the majority rule.

3 Aggregation Games

In this section we present the model of a strategic game played by voters involved in a collective decision-making problem on binary issues. The players' strategies consist of all binary ballots and players' preferences are expressed in the form of a goal that is interpreted on the outcomes of the aggregation (i.e., the collective decision), and by an explicit payoff function for each player i, yielding to i a real number at each profile and encoding, intuitively, the material value he would receive, should that profile of votes occur. Given a set of issues $\mathcal{I}$, let $PS = \{p_1, \ldots, p_m\}$ contain one propositional atom for each issue in $\mathcal{I}$ and $\mathcal{L}_{PS}$ be the propositional language constructed by closing PS under a functionally complete set of Boolean connectives (e.g., $\{\neg, \wedge\}$).

Definition 5 (Aggregation games). *An aggregation game is a tuple $\mathcal{A} = \left\langle \mathcal{N}, \mathcal{I}, F, \{\gamma_i\}_{i\in\mathcal{N}}, \{\pi_i\}_{i\in\mathcal{N}} \right\rangle$ where:*

- *$\langle \mathcal{N}, \mathcal{I} \rangle$ is a binary aggregation structure;*
- *F is an aggregation rule for $\langle \mathcal{N}, \mathcal{I} \rangle$;*
- *each γ_i is a propositional formula in $\mathcal{L}_{PS}$;*
- *$\pi_i : \mathcal{D}^{\mathcal{N}} \to \mathbb{R}$ is a payoff function assigning to each profile a real number denoting the utility of player i.*

A strategy profile in an aggregation game is a profile of binary ballots, and will be denoted with $\boldsymbol{B}$. Intuitively, goals represent positions that players are not willing to sacrifice. When comparing two states, one of which satisfying his goal and one of which not satisfying it, a player will choose the state satisfying his goal. In case of indifference with respect to goals, players will look at the value yielded by the payoff function. This is technically called a *quasi-dichotomous* preference relation [Wooldridge *et al.*, 2013]. Henceforth we employ the logical satisfaction relation $\models$ and its negation $\not\models$.

[2]Cf. the notion of k-resiliency in [Halpern, 2011].

Definition 6. *Let $\mathcal{A} = \left\langle \mathcal{N}, \mathcal{I}, F, \{\gamma_i\}_{i\in\mathcal{N}}, \{\pi_i\}_{i\in\mathcal{N}} \right\rangle$ be an aggregation game, $\boldsymbol{B}, \boldsymbol{B}'$ be two ballot profiles and $i \in \mathcal{N}$ a player. The preference relation $\succeq_i^\pi$ for each $i \in N$ is such that $\boldsymbol{B} \succeq_i^\pi \boldsymbol{B}'$ iff:*

- *$[F(\boldsymbol{B}') \not\models \gamma_i$ and $F(\boldsymbol{B}) \models \gamma_i]$ or*
- *$[F(\boldsymbol{B}') \models \gamma_i \Leftrightarrow F(\boldsymbol{B}) \models \gamma_i]$ and $\pi_i(\boldsymbol{B}) \geq \pi_i'(\boldsymbol{B})$.*

In other words, a profile $\boldsymbol{B}$ is preferred by player i to $\boldsymbol{B}'$ if either $F(\boldsymbol{B})$ satisfies i's goal and $F(\boldsymbol{B}')$ does not or, if both satisfy i's goal or neither do, $\boldsymbol{B}$ yields to i a better payoff than $\boldsymbol{B}'$. Individual preferences over strategy profiles are therefore induced by their goals, by their payoff functions, and by the aggregation procedure used.

A natural class of aggregation games is that of games where the individual utility only depends on the outcome of the collective decision:

Definition 7. *An aggregation game $\mathcal{A}$ is called* uniform *if for all $i \in \mathcal{N}$ and profiles $\boldsymbol{B}$ it is the case that $\pi_i(\boldsymbol{B}) = \pi_i(\boldsymbol{B}')$ whenever $F(\boldsymbol{B}) = F(\boldsymbol{B}')$.*

Games with uniform payoff are arguably the most natural examples of aggregation games. The payoff each player receives is only dependent on the outcome of the vote, and not on the ballot profile that determines it. For convenience, we assume that in uniform games the payoff function is defined directly on outcomes, i.e., $\pi_i : \mathcal{D} \to \mathbb{R}$.

Adapting a standard definition from the literature, we call a strategy B *i-truthful* if it satisfies γ_i. In case γ_i specifies fully a single binary ballot, i.e., the agent has one precise objective over all issues at stake, we fall into the classic setting of having a unique truthful strategy and all other ballots available for strategic voting.

Definition 8. *Let $C \subseteq \mathcal{N}$. We call a strategy profile $\boldsymbol{B} = (B_1, \ldots, B_n)$:*

(i) C-truthful if all B_i with $i \in C$ are i-truthful, i.e., $B_i \models \gamma_i$, for all $i \in C$;

(ii) C-goal-efficient (C-efficient) if $F(\boldsymbol{B}) \models \bigwedge_{i\in C} \gamma_i$;

(iii) totally C-goal-inefficient (totally C-inefficient) if $F(\boldsymbol{B}) \models \bigwedge_{i\in C} \neg\gamma_i$.

One last piece of notation: let us call a game *C-consistent*, for $C \subseteq \mathcal{N}$, if the conjunction of the goals of agents in coalition C is consistent, i.e., if $\bigwedge_{i\in C} \gamma_i$ is satisfiable.

Equilibria in Uniform Aggregation Games

In this section we explore the existence of Nash equilibria (NE) in aggregation games and their properties, paying special attention to NE that are truthful and efficient. We omit the easier proofs in the interest of space.

An aggregation game $\mathcal{A}$ is called *constant*, if all π_i are constant functions, i.e., for all $i \in \mathcal{N}$ and all profiles $\boldsymbol{B}$ we have that $\pi_i(\boldsymbol{B}) = \pi_i(\boldsymbol{B}')$. Recall that a strategy B_i is *weakly dominant* for agent i if for all profiles $\boldsymbol{B}$ we have that $(\boldsymbol{B}_{-i}, B_i) \succeq_i^\pi \boldsymbol{B}$. We start with the following result:

Proposition 1. *If $\mathcal{A}$ is a constant aggregation game for the majority rule, then for every $i \in \mathcal{N}$ every i-truthful strategy is weakly dominant.*

It follows that every constant aggregation game has a NE.

Remark 2 (Generalisation). *Proposition 1 can be generalised to all aggregation rules that are systematic and monotonic, and therefore non-manipulable in social-choice-theoretic sense [Dietrich and List, 2007b]. This moreover shows that with such aggregation procedures dominant strategy equilibria are plenty, namely every profile in which individuals are truthful.*

We now observe that Proposition 1 does not generalise to uniform aggregation games, i.e., games where the utility of players depends solely on the outcome of the aggregation:

Proposition 3. *There exist uniform aggregation games for maj in which truthful strategies are not dominant.*

Proof. Consider the set of issues $\{p, q, t\}$ and a set $\mathcal{N} = \{1, 2, 3\}$. Let moreover $\gamma_i = \top$ for $i = 1, 2$ and $\gamma_3 = t$. Define the payoff function as follows, let $\pi_i(B) = 1$ for $i = 3$ and $B = (0, 1, 0)$, and 0 otherwise. Take the following profiles: $\boldsymbol{B}_1 = ((0, 1, 0), (0, 0, 0), (0, 0, 1))$ and $\boldsymbol{B}_2 = ((0, 1, 0), (0, 0, 0), (0, 1, 0))$. Since $maj(\boldsymbol{B}_1) = (0, 0, 0)$ and $maj(\boldsymbol{B}_2) = (0, 1, 0)$, we have that $\boldsymbol{B}_2 \succ_3^{\pi} \boldsymbol{B}_1$ and that $\boldsymbol{B}_1$, unlike $\boldsymbol{B}_2$, comprises a truthful strategy by 3. □

The fact that truthful voting is not always a dominant strategy for aggregation games might seem counterintuitive, especially when the payoff is required to be uniform across profiles leading to the same outcome. It is however sufficient to recall that when a player is in the position of changing the outcome of the decision in a certain profile – because of being, for example, a veto player – this does not necessarily mean he has the power to satisfy his goal, but he might simply choose the outcome he prefers because of the payoff.

Despite the negative result in Proposition 3, we can still prove the existence of truthful and efficient equilibria in a uniform aggregation game if we assume the mutual consistency of the individual goals of a resilient winning coalition.

Proposition 4. *Let F be a systematic and monotonic rule, and let $C \in \mathcal{W}_F^+$. Every C-consistent uniform aggregation game for F has a NE that is C-truthful and C-efficient.*

On the other hand, undesirable equilibria may occur even when *all* agents have compatible goals.

Proposition 5. *There exist $\mathcal{N}$-consistent aggregation games for maj with NE that are $\mathcal{N}$-truthful and totally $\mathcal{N}$-inefficient.*

Proof. Let $\mathcal{A}$ be an aggregation game for maj such that $\gamma_i = p_i$, and let $\boldsymbol{B}^*$ be the profile illustrated in Table 2. Let all payoff functions π_i be constant. We can observe that $\boldsymbol{B}^*$ is a truthful profile, and therefore it is a NE by Proposition 1 and by the fact that the game is constant. However, the outcome of the majority rule in $\boldsymbol{B}^*$ is totally inefficient, since none of the individual goals are satisfied. □

The goal of Section 4 is to show how to avoid such undesirable equilibria by allowing a pre-vote negotiation phase. We anticipate this by showing the effect of payoff redistributions on the equilibria of the aggregation game.

	W	F	P
Party A	1	0	0
Party B	0	1	0
Party C	0	0	1
Majority	0	0	0

Table 2: Inefficient equilibria

Goal-inefficiency and payoff transformations

We show that goal-inefficiency at equilibrium in uniform aggregation games can be ruled out by means of a redistribution of payoff among the members of a winning coalition. This result is the stepping stone for the framework of Section 4.

Proposition 6. *Let $\mathcal{A} = \langle \mathcal{N}, \mathcal{I}, F, \{\gamma_i\}_{i\in\mathcal{N}}, \{\pi_i\}_{i\in\mathcal{N}} \rangle$ be a uniform aggregation game for a systematic and monotonic procedure F, and let C be a winning coalition for F. Then, there exist payoff functions $\{\pi'_i\}_{i\in C}$ such that $\sum_{i\in C} \pi'_i(\boldsymbol{B}) = \sum_{i\in C} \pi_i(\boldsymbol{B})$ for every profile $\boldsymbol{B}$, and such that the game $\mathcal{A} = \left\langle \mathcal{N}, \mathcal{I}, F, \{\gamma_i\}_{i\in\mathcal{N}}, \{\pi'_i\}_{i\in C} \cup \{\pi_i\}_{i\notin C} \right\rangle$ has no C-inefficient NE.*

Proof. Let B^* be a ballot such that $B^* \models \bigwedge_{i\in C} \gamma_i$. We now construct a redistribution of payoffs in which player $1 \in C$ gives all other players in C an incentive to play B^*, turning it into a weakly dominant strategy. Let $M - 1$ be the maximal payoff difference that some player can obtain between two outcomes in the game $\langle \mathcal{A}, \{\pi_i\}_{i\in\mathcal{N}} \rangle$. The desired payoff functions are constructed as follows. For all $j \neq 1$ such that $j \in C$ define $\pi'_j(\boldsymbol{B}) = \pi_j(\boldsymbol{B}) + M$ for all profiles $\boldsymbol{B}$ with $B_j = B^*$, and $\pi'_j(\boldsymbol{B}) = \pi_j(\boldsymbol{B})$ otherwise. Let finally $\pi'_1(\boldsymbol{B}) = \pi_1(\boldsymbol{B}) - (\sum_{1\neq k\in C} \pi'_k(\boldsymbol{B}) - \sum_{1\neq k\in C} \pi_k(\boldsymbol{B}))$. Observe that the construction of π' ensures that $\sum_{i\in C} \pi'_i(\boldsymbol{B}) = \sum_{i\in C} \pi_i(\boldsymbol{B})$, for every profile $\boldsymbol{B}$. Now let $\overline{\boldsymbol{B}}$ be a C-inefficient profile of the new game and assume towards a contradiction that it is a NE. Take an arbitrary player $j \in C$ such that $\overline{B}_j \neq B^*$. Such a player exists since, by monotonicity of F, if $\overline{B}_i = B^*$ for all $i \in C$ then profile $\overline{\boldsymbol{B}}$ is not C-inefficient. By construction of π'_j and the fact that C is a winning coalition, player $j \in C$ has an incentive to deviate to B^*, hence $\overline{\boldsymbol{B}}$ is not a NE. Contradiction. □

In other words, given a uniform game, payoff functions always exist that can eliminate any NE which is goal-inefficient for a resilient winning coalition, while keeping the sum of players' payoffs constant in the coalition. The new payoff function – which, note, is not necessarily uniform any more – can be thought of as *a binding offer of payoff* that a player makes to the others, incentivising them to deviate to an outcome which is goal-efficient for the coalition.

4 Endogenous Aggregation Games

The games we are going to study have two phases:

- A *pre-vote phase*, where, starting from a uniform aggregation game, players make simultaneous transfers of payoff to their fellow players;

- A *vote phase*, where players play the original aggregation game, but where payoffs are updated according to the transfers occurred in the pre-vote phase.

We call these games *endogenous aggregation games*. The key concept to define them is the one of *transfer function* $\tau_i : \mathcal{D}^{\mathcal{N}} \times \mathcal{N} \to \mathbb{R}_+$ (with $i \in \mathcal{N}$). These functions encode the amount of payoff that a player i gives to player j should a certain profile of votes $\boldsymbol{B}$ be played, in symbols, $\tau_i(\boldsymbol{B}, j)$. We call $\tau \in \prod_i T_i$ a *transfer profile*, denoting by τ^0 the *void* transfer where at every profile every player gives 0 to the others. So by $\tau(\mathcal{A}) = \langle \mathcal{N}, \mathcal{I}, F, \{\gamma_i\}_{i\in\mathcal{N}}, \{\pi_i'\}_{i\in\mathcal{N}} \rangle$ we denote the aggregation game with payoff obtained from $\mathcal{A}$ where π_i' is updated according to the transfer profile τ as follows:

$$\pi_i'(\boldsymbol{B}) = \pi_i(\boldsymbol{B}) + \sum_{j\in\mathcal{N}} \tau_j(\boldsymbol{B}, i) - \sum_{j\in\mathcal{N}} \tau_i(\boldsymbol{B}, j) \qquad (1)$$

It is important to notice that transfers do not preserve the uniformity of payoffs. We define now endogenous aggregation games formally as follows:

Definition 9. *An endogenous aggregation game is a tuple $\mathcal{A}^T = \langle \mathcal{A}, \{T_i\}_{i\in N} \rangle$ where $\mathcal{A}$ is a uniform aggregation game, and each T_i is the set of all functions $\tau_i : \mathcal{D}^{\mathcal{N}} \times \mathcal{N} \to \mathbb{R}_+$.*

Endogenous aggregation games will be analysed as extensive form games with perfect information and simultaneous choices. There are two phases in the overall game and therefore two choice points: in the first one a transfer profile is determined; in the second one a ballot profile is determined. So strategies of player i consist of sequential choices of a transfer τ_i and a ballot B_i. Preference relations are naturally defined between tuples of the form $(\tau, \boldsymbol{B})$: for two profiles of ballots $\boldsymbol{B}, \boldsymbol{B}'$, the expression $(\tau, \boldsymbol{B}) \succeq_i (\tau', \boldsymbol{B}')$ denotes that player i prefers $\boldsymbol{B}$ after τ has been played in the pre-vote phase, to $\boldsymbol{B}'$ after τ' has been played in the pre-vote phase. Equilibrium analysis will be carried out using the solution concept of reference for extensive games: subgame perfect Nash equilibrium (SPE). To avoid complicating the framework with the introduction of mixed strategies over lexicographic preferences – which bears notorious difficulties in terms of game-theoretic analysis [Rubinstein, 2012] – and to keep the focus on pure strategy equilibria, we make the following technical assumption: for any game $\tau(\mathcal{A})$ with some pure strategy NE and any player $i \in N$, *a deviation by i to a game $(\tau_i', \tau_{-i})(\mathcal{A}')$ with no pure strategy NE is never profitable for i.*

Definition 10. *Given an endogenous aggregation game $\langle \mathcal{A}, \{T_i\}_{i\in N} \rangle$, we call a NE $\boldsymbol{B}$ of the aggregation game $\mathcal{A}$ a* surviving Nash equilibrium *(SNE) if there exists a transfer profile τ and a SPE of $\mathcal{A}^T$ where $(\tau, \boldsymbol{B})$ is played on the equilibrium path.*

SPEs can be constructed through backward induction: first, a NE is selected (whenever it exists) after each transfer profile; second, a transfer profile is selected, such that no profitable deviation exists for any player by changing her individual transfer function. Intuitively, surviving Nash equilibria identify those electoral outcomes that can be rationally sustained by an appropriate pre-vote negotiation. Clearly, not all Nash equilibria of the initial game will be surviving equilibria. In what follows we show that surviving equilibria display desirable properties, and pre-play negotiations can effectively act as equilibrium refinement tools for aggregation games.

Equilibria in Endogenous Aggregation Games

Pre-vote negotiations can be shown to yield desirable effects in terms of goal-efficiency, as shown in the following:

Theorem 7. *Let $\mathcal{A}^T = \langle \mathcal{A}, \{T_i\}_{i\in N} \rangle$ be an endogenous aggregation game for a systematic and monotonic aggregator F and let $C \in \mathcal{W}_F^+$. Then, every C-efficient NE of $\langle \mathcal{A}, \{\pi_i\}_{i\in\mathcal{N}} \rangle$ is a surviving NE.*

Proof. Let $\boldsymbol{B}$ be a C-efficient NE of $\mathcal{A}$. There exists at least one by Proposition 4 and by the fact that $\mathcal{A}$ is uniform. We want to find a transfer function τ^* such that $(\tau^*, \boldsymbol{B})$ is a SPE of $\mathcal{A}^T$. Let $M - 1$ be the maximal payoff difference between outcomes as defined in the proof of Proposition 6. For all $i, j \in N$, let $\tau_i^*(\boldsymbol{B}', j) = 2M$ if $B_i' \neq B_i$, and $\tau_i^*(\boldsymbol{B}', j) = 0$ otherwise. In words, each player i is committing to play the ballot B_i by offering the others $2M$ in case of deviation. Now we show that $(\tau^*, \boldsymbol{B})$ is a SPE using the fact that $\boldsymbol{B}$ is C-efficient (i.e., all deviations are payoff based) and generalising the argument for pre-play negotiations with more than two players given in the proof of Jackson and Wilkie [2005, Theorem 4] to our case, as follows. First, notice that after the choice of τ^*, B_i is (uniquely) surviving iterated elimination of strictly dominated strategies for each player i. Crucially, by the fact that $C \in \mathcal{W}_F^+$, it is so for players in $\mathcal{N} \backslash C$. Set $(\tau^*, \boldsymbol{B})$ to be the choices on the equilibrium path, while off the equilibrium path select any Nash Equilibrium, whenever there is one. We show that a deviation to some τ_{i^*}'' by a player i^* is not profitable for i^*. Such a deviation can only be improving if it leads to play something other than B_j by some other player j. If only i^* changed actions, then i^* cannot do better given that $\boldsymbol{B}$ was a Nash equilibrium and $\tau_{i^*}^*(\boldsymbol{B}', j) = 0$ when $B_i' = B_i$. Now consider the case where a pure strategy Nash equilibrium $\boldsymbol{B}''$ is played in the second stage where $B_j'' \neq B_j$ for some $j \neq i^*$. Let there be $k \geq 1$ players $j \neq i^*$ for which $B_j'' \neq B_j$ and consider some such j. By playing B_j'' player j's payoff is $\pi_j(\boldsymbol{B}'') - (|N| - 1)2M + 2M(k+1) + \tau_{i^*}''(\boldsymbol{B}'', j)$. If j plays B_j instead, then j's payoff is $\pi_j(B_j, \boldsymbol{B}_{-j}'') + 2M(k+1) + \tau_{i^*}''(B_j, \boldsymbol{B}_{-j}'', j)$. As $\boldsymbol{B}''$ is a NE, it must be the case that $\tau_{i^*}''(\boldsymbol{B}'', j) - \tau_{i^*}''(B_j, \boldsymbol{B}_{-j}'', j) \geq \pi_j(B_j, B_{-j}'') - \pi_j(\boldsymbol{B}'') + (|\mathcal{N}| - 1)2M$. Given the definition of M and given the fact that $|\mathcal{N}| - 1 \geq 2$ it follows that $\tau_{i^*}''(\boldsymbol{B}'', j) - \tau_{i^*}''(B_j, \boldsymbol{B}_{-j}'', j) > 3M$, which, by the definition of transfer function, implies that $\tau_{i^*}''(\boldsymbol{B}'', j) > 3M$. Therefore i^*'s utility in the new equilibrium is at most $\pi_{i^*}(\boldsymbol{B}'') - k3M + k2M$. The fact that $k \geq 1$ implies that $\pi_{i^*}(\boldsymbol{B}'') - k3M + k2M \leq \pi_{i^*}(\boldsymbol{B})$. Because of C-efficiency of $\boldsymbol{B}$, monotonicity of F, and the fact that $C \in \mathcal{W}_F^+$, we have that the constructed deviation cannot be profitable. $\square$

The converse of Theorem 7 holds true, provided that the goals of a winning coalition are mutually consistent:

Theorem 8. *Let $\mathcal{A}^T = \langle \mathcal{A}, \{T_i\}_{i\in N} \rangle$ be an endogenous aggregation game for a systematic and monotonic aggregator*

F such that $\mathcal{A}$ is C-consistent for $C \in \mathcal{W}_F$. Then, every surviving NE of $\mathcal{A}^T$ is C-efficient.

Proof. We proceed by contraposition. Let $\boldsymbol{B}^*$ be a NE that is not C-efficient, i.e., such that $F(\boldsymbol{B}^*) \not\models \gamma_i$ for some individual $i \in C$, and assume that $\boldsymbol{B}^*$ is a SNE. Therefore there exists a transfer function τ^* and a SPE of $\mathcal{A}^T$ such that $(\tau^*, \boldsymbol{B}^*)$ is played on the equilibrium path. We now construct a profitable deviation from τ^*, leading to contradiction. By C-consistency of $\mathcal{A}$ there exists a ballot B' such that $B' \models \bigwedge_{j \in C} \gamma_j$, hence in particular $B' \models \gamma_i$. Let now i deviate to any transfer profile $\tau' = (\tau'_i, \tau^*_{-i})$ such that she offers more than the payoff difference to all other players if they vote for ballot B', i.e., $\tau'_i(B'_j, \boldsymbol{B}''_{-j}, j) > \pi_j(\boldsymbol{B}'') + \tau^*_i(\boldsymbol{B}'', j) - \pi_j(B'_j, \boldsymbol{B}''_{-j})$, for each $j \in N$, and each $\boldsymbol{B}''_{-j}$. By the fact that B' is C-efficient, F systematic and monotonic, and C is a winning coalition, this transfer makes each B'_j, with $j \in N \setminus \{i\}$, (uniquely) survive iterated elimination of strictly dominated strategies. The worst for i NE will now satisfy γ_i, making τ' a profitable deviation. □

These results suggest that pre-vote negotiations are a powerful tool players have to overcome the inefficiencies of aggregation rules. More specifically, when the goals of players of some resilient coalition can be satisfied, pre-vote negotiations allow players to engineer side-payments leading to equilibrium outcomes that satisfy them, ruling out all the others. We stress that players' equilibrium strategies in the two phase game *remain individually rational strategies* and the game remains non-cooperative throughout – and hence radically different from approaches like [Bachrach *et al.*, 2011] – even when equilibrium strategies end up sustaining efficiency.

Remark 9 (Algorithms). *An algorithm to compute a pre-vote negotiation strategy that leads to a sustainable NE is provided in the proofs of Theorems 7 and 8. The assumption of perfect information is crucial here, and can be considered as an approximation of a real-world situation in which the goals and payoffs of the agents can be assessed by means, e.g., of a poll.*

Pre-vote negotiations and voting paradoxes

We show an application of endogenous aggregation games to binary aggregation with constraints, or judgment aggregation [Grandi and Endriss, 2013; Grossi and Pigozzi, 2014], where individual ballots need to satisfy a logical formula, the *integrity constraint*, to be considered *admissible*. In case each individual provides an admissible ballot, the obvious question is whether the outcome of a given aggregation rule will be admissible, as well. Here is an instance of this problem.

Example 2. *Consider the scenario in Table 1. In line with the intuitions behind the example, we stipulate that accepting W while at the same time rejecting both F and P is not an admissible opinion: if one wants to develop atomic weapons one should either import nuclear technology or develop it domestically. We can formulate this requirement in a propositional language as $W \to (F \vee P)$, making ballot $(1,0,0)$ inadmissible. All submitted ballots in the example satisfy this requirement but the majority ballot does not (Table 1).*

Paradoxical situations as those in Example 2 can be viewed as undesirable outcomes of aggregation games. Assume to this purpose each party to have the following goals: $\gamma'_A = W, \gamma'_B = F, \gamma'_C = \neg P$. Let $\pi_A = \pi_B = \pi_C$ be constant payoff functions. Observe that parties' goals are all consistent with the integrity constraint $W \to (F \vee P)$, and that admissible ballot (1,1,0) satisfies each of them. The profile in Table 1 shows a truthful NE that however does not satisfy neither the goal of party B nor the integrity constraint $W \to (F \vee P)$. However, this equilibrium is *not* surviving because party B could transfer enough utility to party C for it to vote for F.

In consistent aggregation games equilibria that give rise to a voting paradox may not survive, whereas equilibria avoiding such paradoxes are *always* sustained by a pre-vote negotiation phase. But the key question is whether we can guarantee that inadmissible equilibria do not survive.The following proposition shows a simple sufficient condition. Let the integrity constraint be a formula IC, and call an individual i *responsible* if $\gamma_i \models \mathrm{IC}$, i.e., if i's goal logically implies the constraint. The following holds:

Proposition 10. *If $\langle \mathcal{A}, \{T_i\}_{i \in N} \rangle$ is an $\mathcal{A}^{\mathcal{T}}$-game such that $\mathcal{A}$ is consistent and there exists a responsible player, then every surviving equilibrium is* IC*-consistent.*

In particular, if all individual goals imply the integrity constraint, i.e., the goal of each party includes an admissible decision, pre-vote negotiation will rule out all inadmissible equilibria and some admissible outcome is bound to survive.

5 Acknowledgments

Paolo Turrini acknowledges the support of the Marie Curie fellowship "NINA" (FP7- PEOPLE-2012-IEF, 327424).

6 Conclusions

In this paper we studied the effect of a pre-vote phase before an aggregation game for binary voting, where voters might hold an objective about a subset of the issues at stake while willing to strike deals on the remaining ones. A number of papers in the literature on voting games have focused on the problem of avoiding undesirable equilibria (e.g., [Desmedt and Elkind, 2010] and [Obraztsova *et al.*, 2013]). Our proposal has been to study an explicit pre-vote negotiation phase, during which agents can influence one another before casting their ballots in order to obtain an individually more favourable electoral outcome. By doing so, we have shown how undesirable equilibria can be eliminated (dually, desirable ones sustained) as an effect of a rational distributed negotiation phase, for a set of aggregators defined axiomatically. We have also seen how these results have potential consequences in avoiding paradoxical situations of aggregation procedures, a core research problem in judgment aggregation [Grossi and Pigozzi, 2014].

Future work include the study of aggregation games for different voting procedures, e.g., distance-based ones, that do not satisfy the axiom of systematicity. A second important line of work is the study of budgeted transfer functions, in which agents are endowed with limited resources to be used in the negotiation phase. Finally, as observed in Remark 9, a treatment of imperfect information is naturally called for.

References

[Bachrach *et al.*, 2011] Y. Bachrach, E. Elkind, and P. Faliszewski. Coalitional voting manipulation. In *Proc. of the 22nd International Joint Conference on Artificial Intelligence (IJCAI-2011)*, 2011.

[Baumeister *et al.*, 2013] D. Baumeister, G. Erdélyi, O. Johanna Erdélyi, and J. Rothe. Computational aspects of manipulation and control in judgment aggregation. In *Proc. of the Third International Conference on Algorithmic Decision Theory (ADT-2013)*, 2013.

[Brandt *et al.*, 2014] F. Brandt, V. Conitzer, and U. Endriss. Computational social choice. In G. Weiss, editor, *Multiagent Systems*, pages 213–283. MIT Press, 2014.

[Brânzei *et al.*, 2013] S. Brânzei, I. Caragiannis, J. Morgenstern, and A. Procaccia. How bad is selfish voting? In *Proc. of the 27th Conference on Artificial Intelligence (AAAI-13)*, 2013.

[Bredereck *et al.*, 2014] R. Bredereck, J. Chen, S. Hartung, S. Kratsch, R. Niedermeier, O. Suchý, and G. J. Woeginger. A multivariate complexity analysis of lobbying in multiple referenda. *Journal of Artificial Intelligence Research*, 50:409–446, 2014.

[Christian *et al.*, 2007] R. Christian, M. Fellows, F. Rosamond, and A. Slinko. On complexity of lobbying in multiple referenda. *Review of Economic Design*, 11(3):217–224, 2007.

[Desmedt and Elkind, 2010] Y. Desmedt and E. Elkind. Equilibria of plurality voting with abstentions. In *Prooceedings of the 11th ACM Conference on Electronic Commerce (EC-2010)*, 2010.

[Dhillon and Lockwood, 2004] Amrita Dhillon and Ben Lockwood. When are plurality rule voting games dominance-solvable? *Games and Economic Behavior*, 46(1):55 – 75, 2004.

[Dietrich and List, 2007a] F. Dietrich and C. List. Arrow's theorem in judgment aggregation. *Social Choice and Welfare*, 29(19–33), 2007.

[Dietrich and List, 2007b] F. Dietrich and C. List. Strategy-proof judgment aggregation. *Economics and Philosophy*, 23(3):269–300, 2007.

[Dryzek and List, 2003] J. Dryzek and C. List. Social choice theory and deliberative democracy: A reconciliation. *British Journal of Political Science*, 33:1–28, 2003.

[Grandi and Endriss, 2013] Umberto Grandi and Ulle Endriss. Lifting integrity constraints in binary aggregation. *Artificial Intelligence*, 199-200:45–66, 2013.

[Grossi and Pigozzi, 2014] Davide Grossi and Gabriella Pigozzi. *Judgment Aggregation: A Primer*. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan & Claypool Publishers, 2014.

[Gueth and Selten, 1991] W. Gueth and R. Selten. Majority voting in the condorcet paradox as a problem of equilibrium selection. In *Game Equilibrium Models IV*. 1991.

[Halpern, 2011] J. Halpern. Beyond nash-equilibrium: Solution concepts for the 21st century. In K. Apt and E. Grädel, editors, *Lectures in Game Theory for Computer Scientists*, pages 264–289. Cambridge University Press, 2011.

[Harrenstein *et al.*, 2001] Paul Harrenstein, Wiebe van der Hoek, John-Jules Ch. Meyer, and Cees Witteveen. Boolean games. In *Proceeding of the Eighth Conference on Theoretical Aspects of Rationality and Knowledge (TARK VIII)*, pages 287–298, 2001.

[Hazon *et al.*, 2013] Noam Hazon, Raz Lin, and Sarit Kraus. How to change a group's collective decision? In *Proc. of the 23rd International Joint Conference on Artificial Intelligence (IJCAI-2013)*, 2013.

[Jackson and Wilkie, 2005] Matthew O. Jackson and Simon Wilkie. Endogenous games and mechanisms: Side payments among players. *Review of Economic Studies*, 72(2):543–566, 2005.

[Kim, 1996] Y. Kim. Equilibrium selection in n-person coordination games. *Games and Economic Behavior*, 15:203–227, 1996.

[Lev and Rosenschein, 2012] Omer Lev and Jeffrey S. Rosenschein. Convergence of iterative voting. In *Proc. of the 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS-2012)*, 2012.

[List, 2011] C. List. Group communication and the transformation of judgments: An impossibility result. *The Journal of Political Philosophy*, 19(1):1–27, 2011.

[Meir *et al.*, 2010] Reshef Meir, Maria Polukarov, Jeffrey S. Rosenschein, and Nicholas R. Jennings. Convergence to equilibria in plurality voting. In *Proc. of the 24th conference on Artificial Intelligence (AAAI-2010)*, 2010.

[Messner and Polborn, 2007] Matthias Messner and Mattias Polborn. Strong and coalition-proof political equilibria under plurality and runoff rule. *International Journal of Game Theory*, 35(2):287–314, 2007.

[Meyerson, 1978] R. B. Meyerson. Refinements of the nash equilibrium concept. *International Journal of Game Theory*, 7(2):73–80, 1978.

[Obraztsova *et al.*, 2013] Svetlana Obraztsova, Evangelos Markakis, and David R. M. Thompson. Plurality voting with truth-biased agents. In *Proc. of the 6th Int. Symposium on Algorithmic Game Theory (SAGT-2013)*, 2013.

[Rubinstein, 2012] Ariel Rubinstein. *Lecture Notes in Microeconomic Theory: The Economic Agent*. Princeton University Press, 2012.

[Turrini, 2013] P. Turrini. Endogenous boolean games. In *Proc. of the 23rd International Joint Conference on Artificial Intelligence (IJCAI-2013)*, 2013.

[Wooldridge *et al.*, 2013] M. Wooldridge, U. Endriss, S. Kraus, and J. Lang. Incentive engineering in boolean games. *Artificial Intelligence*, 195:418–439, 2013.

[Xia and Conitzer, 2010] Lirong Xia and Vincent Conitzer. Stackelberg voting games: Computational aspects and paradoxes. In *Proc. of the 24th Conference on Artificial Intelligence (AAAI-2010)*, 2010.