Pre-vote negotiations and binary voting with constraints

Umberto Grandi$^1$, Davide Grossi$^2$, and Paolo Turrini$^3$

$^1$Department of Mathematics, University of Padova
$^2$Department of Computer Science, University of Liverpool
$^3$Department of Computing, Imperial College London

Abstract

We study voting games on possibly interconnected issues, where voters might hold a principled opinion about a subset of the issues at stake while willing to strike deals on the remaining ones, and can influence one another before casting their ballots in order to obtain an individually more favourable outcome. We analyse voters’ rational behaviour in a two-phase game, allowing players to undergo a negotiation phase before their vote, and showing under what conditions undesirable equilibria can be removed as an effect of the pre-vote phase.

1 Introduction

Collective decision-making often requires the ability of compromising and striking deals. It is therefore a dynamic process rather than a static one, as too often depicted in the literature on social choice theory. Being able to formally describe the process leading to an agreement facilitates the implementation of systems that can reach human-level collective decisions and the analysis of human decision-making.

Voting is a collective decision procedure where persuasion can be an effective tool to induce opinion change and settle for compromise solutions [1, 2]. Indeed, in strategic interaction pre-play negotiations are known to be effective in overcoming inefficient allocations caused by players’ individual rationality [3]. When players are allowed to offer a part of their gains at certain outcomes to influence the decisions of their opponents and have enough resources for doing so, they can even overcome otherwise highly inefficient scenarios, such as the Prisoners’ Dilemma [3].

But voting is a specific type of strategic interaction, where individuals may hold uncompromising opinions about certain issues at stake – think of discussions on nuclear energy policies – and be therefore hard to persuade to change their behaviour. Moreover, collective decisions are determined by voting rules.
– e.g., unanimous, majoritarian etc. –, which directly constrain the amount of compromise needed to reach consensus. In these situations, overcoming paradoxical consequences typical of aggregation procedures [4], crucially depends on how much, if anything at all, voters are willing to give up to accommodate the others.

In this paper we study pre-vote negotiations in voting games over logically interrelated issues, where voters hold a special type of lexicographic preferences over the set of issues at stake, i.e., hold a principled opinion about a subset of them while willing to negotiate on the remaining ones, and can influence one another before casting their ballots by means of monetary incentive in order to obtain an individually more favourable outcome. The model will be abstract enough to generalise a variety of different pre-vote opinion change phenomena such as: deliberation, negotiation, power and authority, bribery, etc. All these phenomena will be studied as multi-agent processes whereby voters invest resources (e.g., persuasiveness, money, time, personal credibility, etc.) to influence one another.

Scientific context Our approach stems from a variety of existing lines of research in social choice and game theory. Specifically, we build up on three of them:

Judgment Aggregation. We study societies of voters that express a yes/no opinion on several issues at stake and whose ballots need to stick to a (possibly empty) set of constraints [5]. The strategic issues involved in this type of voting setting have, to the best of our knowledge, never been explored and will be our focus here. The approach we take is related to the growing literature on voting games. Classical references include the work of Dhillon and Lockwood [6] and Messner and Polborn [7].

Boolean games. We model decision strategies in judgment aggregation as Boolean Games [8, 9], 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 action. The dichotomy payoff-goals, typical of boolean games, allows us to encode a truly lexicographic preference relation, i.e., with goals representing outcomes satisfying non-negotiable standpoints and numerical values modelling secondary (monetary) aspects.

Pre-play negotiations. We model pre-vote negotiations in Boolean Games in the framework of Endogenous Boolean Games [10], which studies boolean games together with a pre-play phase. In the spirit of Jackson and Wilkie [3] players are entitled to sacrifice a part of the payoff received at a certain outcome to convince their opponents to play certain strategies, which in our case consist of voting ballots.

Paper contribution and outline The type of issues we can address in this framework concern the effect of pre-play negotiations on the outcome of col-
lective decisions, modelled as pre-vote phase in a boolean game for judgment aggregation. In particular we will be able to show under what conditions bad equilibrium outcomes – e.g., inefficient ones – can be overcome, and dually good ones sustained, thanks to a pre-play phase. A typical showcase will be those games guaranteeing goal-efficient profiles (the ones satisfying the goals of all players) after pre-play negotiation.

We present our framework in three steps. First, we define voting games for judgment aggregation using purely dichotomous preferences, without explicit payoffs, analyzing the properties of the equilibria (Section 3). Second, we add an explicit payoff function to the previously defined structures, moving to a proper lexicographic (quasi-dichotomous) preference relation, showing how undesirable equilibria can be removed by appropriate modifications of the game matrix (Section 4). 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 undesirable equilibria can be removed – and, dually, desirable ones kept – by means of rational negotiations (Section 5).

Even though only the last step is dedicated to the actual effect of pre-vote negotiations, the first two illustrate the properties of collective decisions in their absence, and are therefore necessary in order to understand the added value of having a pre-vote phase.

2 Preliminaries

We model situations of collective decision-making using a general framework for judgment aggregation called binary aggregation with integrity constraints [5]. In this setting a finite set of individuals express yes/no opinions on a finite set of binary issues, and these opinions are then aggregated into a collective decision over each issue. This framework has been shown to be general enough to account for classical settings such as preference aggregation and judgment aggregation, and to be particularly effective in the study of paradoxes of collective decision-making.

**Definition 1 (BA structure).** Let $PS = \{p_1, \ldots, p_m\}$ be a set of propositional atoms and $L_{PS}$ be the propositional language constructed by closing $PS$ under standard propositional connectives. A binary aggregation structure (BA structure) is a tuple $S = (N, I, IC)$ where:

- $N = \{1, \ldots, n\}$ is a finite set of individuals;
- $I = \{1, \ldots, m\}$ is a finite set of issues;
- $IC$ is a propositional formula of $L_{PS}$.

We denote $D = \{B \mid B : I \rightarrow \{0, 1\}\}$ the set of all possible binary opinions over the set of issues $I$ and call an element $B \in D$ a ballot. The integrity constraint $IC$ is a propositional formula, e.g., $p_1 \land p_2 \rightarrow p_3$, that can be interpreted on elements of $D$ and can be thus used to define the set of admissible ballots as the set $\text{Mod}(IC) = \{B \in D \mid B \models IC\}$. 
\[ IC = (W \land \neg F) \rightarrow P \]

|       | 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 the Discursive Dilemma

**Example 1.** A parliament composed by equally representative parties A, B, C is to decide whether to build nuclear power plants (P) and develop atomic weapons (W). If importing nuclear technology from abroad is not an option, the development of atomic weapons involves the construction of nuclear power plants. We can model this situation as an integrity constraint \[ IC = (W \rightarrow P) \], making ballot \((1, 0)\) – in words, voting positively for \(W\) and negatively for \(P\) – an inadmissible ballot.

An admissible profile \(B = (B_1, \ldots, B_n)\) is the choice of an admissible ballot for every individual in \(N\). We write \(B_i\) to denote the ballot of individual \(i\) within a profile \(B\). Thus, \(B_i(j) = 1\) indicates that individual \(i\) accepts issue \(j\) in profile \(B\). Furthermore we denote by \(N_j^B = \{ i \in N \mid B_i(j) = 1 \}\) the set of individuals accepting issue \(j\) in profile \(B\).

**Definition 2 (Aggregation procedure).** Given a BA structure \(S\), an aggregation procedure for \(S\) is a function \(F : \text{Mod}(\text{IC})^N \rightarrow D\), mapping every profile of admissible ballots to a binary ballot in \(D\). We denote with \(F(B)_j\) the outcome of the aggregation on issue \(j\).

Examples of aggregation functions include the issue-by-issue majority rule \((\text{maj})\) which accepts an issue if and only if there is a majority of voters accepting it.

Aggregation rules are classified by means of axioms that bind the properties of the outcome in certain profiles. A prime example is the axiom of unanimity, which demands that the outcome of aggregation coincide with the individuals’ judgments in case of consensus. In the remainder of the paper we shall consider aggregation games with the majority rule, however many of our results can be generalised to arbitrary aggregation procedures that satisfy a number of relevant axioms. Due to space restrictions, we refer the reader to the relevant literature for a formal treatment of axiomatic properties \([5]\).

Given a BA structure \(S\), an aggregation procedure \(F\) is said to be collectively rational with respect to \(S\) if the outcome \(F(B)\) is an admissible ballot for every admissible profile \(B\). The profile in Table 1 shows that the majority rule is not always collectively admissible.
3 Aggregation Games

In this section we present the simplest model of strategic reasoning by players involved in a collective decision-making problem on binary issues. The players’ strategies consist of all admissible 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).

**Definition 3 (Aggregation games).** An aggregation game is a tuple $A = \langle N, I, IC, F, \{\gamma_i\}_{i \in N} \rangle$ such that:

- $\langle N, I, IC \rangle$ is a binary aggregation structure;
- $F$ is an aggregation procedure for $\langle N, I, IC \rangle$;
- each $\gamma_i$ is a propositional formula in $L_{PS}$ which is consistent with $IC$;

All individuals share the same set of admissible strategies, namely the set of admissible ballots $Mod(IC)$. A strategy profile is therefore a profile of (admissible) binary ballots, and will be denoted with $B$. Preferences in aggregation games are only formulated in terms of goal satisfaction.

**Definition 4 (Preferences in aggregation games).** Let $A = \langle N, I, IC, F, \{\gamma_i\}_{i \in N} \rangle$ be an aggregation game. For all profiles $B, B'$ and $i \in N$, the preference relation $\succeq_A^i$ is such that $B \succeq_A^i B' \iff F(B') \not\models \gamma_i$ or $F(B) \models \gamma_i$.

We call a strategy $B \in Mod(IC)$ truthful for agent $i$ if it satisfies $\gamma_i$. We call a game consistent if the conjunction of all individual goals $\bigwedge_i \gamma_i$ is consistent with IC.

**Definition 5.** We call a strategy profile $B = (B_1, \ldots, B_n)$:

- truthful if all $B_i$ are truthful;
- $IC$-consistent if $F(B) \models IC$;
- goal-efficient if $F(B) \models \bigwedge_i \gamma_i$;
- goal-inefficient if $F(B) \not\models \gamma_i$ for all $i \in N$.

We need one last definition specifying a condition under which the goals of the voters are logically independent from one another:

**Definition 6.** If $A$ is a set of propositional variables, let $L_A$ be the propositional language constructed over $A$. An aggregation game is said to have partitioned goals if there is a partition of issues $I = A_1 \cup \cdots \cup A_n$ such that $\gamma'_i \in L_{A_i}$ for some $\gamma'_i$ with $B \models \gamma_i \leftrightarrow \gamma'_i$, for all $B \in Mod(IC)$. 


3.1 Equilibria in 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, goal-efficient and IC-consistent. We start with the following proposition:

**Proposition 1.** If $\mathcal{A}$ is an aggregation game for maj then every truthful strategy for $i$ is (weakly) dominant.

**Proof.** Let $B^*_i$ be a truthful strategy for agent $i$ in $\mathcal{A}$ and let $B'_i$ any non-truthful strategy for $i$. WLOG we want to show that for each strategy profile $B_{-i}$ we have that $(B_{-i}, B^*_i) \succeq_A (B_{-i}, B'_i)$. We have four cases. Both $\text{maj}(B_{-i}, B^*_i)$ and $\text{maj}(B_{-i}, B'_i)$ satisfy $\gamma_i$ or do not satisfy it. In both these cases $B^*_i$ weakly dominates $B'_i$. If $\text{maj}(B_{-i}, B^*_i)$ satisfies $\gamma_i$ and $\text{maj}(B_{-i}, B'_i)$, then $B^*_i$ strictly dominates $B'_i$. Finally, if $\text{maj}(B_{-i}, B'_i)$ satisfies $\gamma_i$ but $\text{maj}(B_{-i}, B^*_i)$ does not, then the majority for $\gamma_i$ is larger in $(B_{-i}, B'_i)$ then in $(B_{-i}, B^*_i)$, against the assumption that $B^*_i$ is truthful. Hence $B^*_i$ is weakly dominant.

**Corollary 2.** Every aggregation game for the majority rule has a pure strategy Nash equilibrium.

Proposition 1 and thus Corollary 2 as well, can be generalised to all aggregation procedures that are independent and monotonic (therefore non-manipulable). This moreover shows that with such aggregation procedures dominant strategy equilibria are plenty, namely every profile in which individuals are truthful. It must be recalled that we do not require the conjunction of goals to be consistent with IC, which means that there may be truthful dominant equilibria leading to outcomes that are not admissible (IC-inconsistent).

In what follows we show a positive result stating the existence of truthful and goal-efficient equilibria in any aggregation game for the majority rule in which goals are consistent:

**Proposition 3.** Every consistent aggregation game for maj has an IC-consistent NE that is truthful and goal-efficient.

**Proof.** Since $\mathcal{A}$ is consistent, then there exists an admissible ballot $B^*$ such that $B^* \models \bigwedge_i \gamma_i$. Let $B^* = (B^*_1, \ldots, B^*_n)$, i.e., $B^*$ is a unanimous profile where each individual votes for $B^*$. Since the profile is unanimous, $\text{maj}(B^*) = B^*$. Clearly $\text{maj}(B^*)$ satisfies $\bigwedge_i \gamma_i$ and IC, and each individual votes truthfully. Hence $B^*$ is a NE by Proposition 1.

Proposition 3 has a negative counterpart: the existence of consistent aggregation games with undesirable equilibria.

**Proposition 4.** There exist consistent aggregation games for maj with an IC-consistent NE that is truthful and goal-inefficient, and consistent aggregation games for maj with a truthful and goal-inefficient NE that is IC-inconsistent.
Table 2: IC-inefficient equilibria

| Voter | p1 | p2 | p3 |
|-------|----|----|----|
| Voter 1 | 1  | 0  | 0  |
| Voter 2 | 0  | 1  | 0  |
| Voter 3 | 0  | 0  | 1  |
| Majority | 0  | 0  | 0  |

Proof. For the first statement, let \( A \) be an aggregation game for \( \text{maj} \) such that \( \gamma_i = p_i \), let \( IC = \top \) and let \( B^* \) be the profile illustrated in Table 2. We can observe that \( B^* \) is a truthful profile – therefore a NE by Proposition 1 – for which however the outcome of the aggregation satisfies no individual goals, although satisfying IC. For the second statement, consider a different game \( A' \) with the same individual goals but \( IC' = p_1 \lor p_2 \lor p_3 \). Then \( B^* \) is a truthful NE which is both goal-inefficient and IC-inconsistent.

Undesirable equilibria are not always present in every aggregation game, unlike desirable ones, but in some particular cases we have a guarantee of their existence.

Proposition 5. The following hold:

- Every consistent game with \( IC = \top \) and partitioned goals has a NE which is IC-consistent, truthful and goal-inefficient.
- Every consistent game such that \( IC \models \lor_{i \in N} \gamma_i \) and with partitioned goals has a NE which is IC-inconsistent, truthful and goal-inefficient.

Proof. The statements follow from a straightforward generalization of the scenario illustrated in Table 2.

Example 2. Consider the judgment aggregation structure of the discursive dilemma as presented in Table 1 and assume each party has goals equivalent to a full admissible ballot:

\[
\begin{align*}
\gamma_1 &= W \land \neg F \land P \\
\gamma_2 &= W \land F \land \neg P \\
\gamma_3 &= \neg W \land \neg F \land \neg P
\end{align*}
\]

The strategies available to player \( i \) are all valuations for the three issues \( W, F \) and \( P \). There is however only one truthful strategy for each voter, which is weakly dominant, and which corresponds to her goal. The discursive dilemma profile is therefore a dominant strategy equilibrium – hence a NE – which is truthful, goal-inefficient and IC-inconsistent.
3.2 Boolean Games and Voting Games

Each aggregation game can be seen as a special kind of boolean game \[8\]. Formally, given an aggregation game \(A = \langle N, I, IC, F, \{\gamma_i\}_{i \in N} \rangle\), we can define the corresponding boolean game \(B_A = \langle N_A, P_A, \{P_A^i\}_{i \in N_A}, \{g_A^i\}_{i \in N_A} \rangle\) as follows: Take the same set of individuals \(N_A = N\) and as propositional variables the set \(P_A = PS \times N\). Each individuals can control the set of variables \(P_A^i = \{(p, i) \mid p \in PS\}\), generating a partition \(\{P_A^i\}_{i \in N_A}\) of \(P_A\). As for the goals, let \(g_A^i\) be a formula in \(LP_A\) such that for every valuation \(V : P_A \rightarrow \{0,1\}\) we have that \(V \models g_A^i\) if and only if \(F(V_1, \ldots, V_n) \models \gamma_i\), where \(V_i = V(-, i)\) is the binary ballot obtained from \(V\) by setting the value of the second argument to \(i\).

Since binary voting with constraints generalizes preference aggregation settings, where the outcome of voting is a collective preference order among candidates/alternatives \([5]\), aggregation games can be viewed as a generalization of voting games, where goals represent constraints each voter would like the collective preference order to satisfy (e.g., that candidate \(a\), even if not at the top of the order, is ranked above candidate \(b\)).

4 Aggregation Games with Payoff

This section refines the model of aggregation games, introducing an explicit payoff function for each player \(i\), yielding to \(i\) a real number at each profile and encoding, intuitively, the amount of resources that he would receive, should that profile of votes occur.

Definition 7 (A\(^\pi\) games). An aggregation game with payoff is a tuple \(\langle A, \{\pi_i\}_{i \in N} \rangle\) where \(A\) is an aggregation game and \(\pi_i : Mod(IC)^N \rightarrow \mathbb{R}\) is a payoff function assigning to each admissible profile a real number denoting the utility of player \(i\).

Intuitively, goals represent uncompromising positions upon which players are not willing to negotiate. Still, players are able to compare any two states, which both satisfy or both falsify their own goals, by looking at the value yielded by the payoff function. This is technically called a quasi-dichotomous preference relation \([9]\):

Definition 8 (Preferences in A\(^\pi\) games). Let \(\langle A, \{\pi_i\}_{i \in N} \rangle\) be an A\(^\pi\)-game, \(B, B'\) be two ballot profiles and \(i \in N\) a player. The preference relation \(\geq_i^\pi\) for each \(i \in N\) is such that \(B \geq_i^\pi B'\) iff:

- \([F(B') \not\models \gamma_i \land F(B) \models \gamma_i]\) or
- \([F(B') \models \gamma_i \Leftrightarrow F(B) \models \gamma_i\) and \(\pi_i(B) \geq \pi_i'(B))\).

In other words, a profile \(B\) is preferred by player \(i\) to \(B'\) if either \(F(B)\) satisfies \(i\)'s goal and \(B'\) doesn’t or, if both satisfy \(i\)'s goal or neither do, \(F(B)\) yields to \(i\) a better payoff than \(F(B')\).
A natural class of $A^\pi$-games is that of games where the individual utility only depends on the outcome of the collective decision, that is, for all $i \in \mathcal{N}$, $\pi_i(B) = \pi_i(B')$ whenever $F(B) = F(B')$. We call these games uniform.

Games with uniform payoff are the simplest generalisation of aggregation games, in which the preference defined by the individual goals is refined by introducing a payoff function that is defined on the outcome of the aggregation. For convenience, we assume that in uniform $A^\pi$-games the payoff function is defined directly on outcomes, i.e., $\pi_i : D \rightarrow \mathbb{R}$.

The definition of truthful strategies, as well as Definition 5, still apply to this setting.

### 4.1 Equilibria in games with uniform payoff

The set of NE for a uniform $A^\pi$-game $\langle A, \{\pi_i\}_{i \in \mathcal{N}} \rangle$ is a subset of the NE of $A$. This is a straightforward consequence of our definition of quasi-dichotomous preferences. Moreover, uniform $A^\pi$-games for the majority rule do not necessarily have dominant truthful strategies:

**Proposition 6.** There exist $A^\pi$-games for maj in which truthful strategies are not dominant.

*Proof.\footnote{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: $B_1 = ((0, 1, 0), (0, 0, 0), (0, 0, 1))$ and $B_2 = ((0, 1, 0), (0, 0, 0), (0, 1, 0))$. Since $maj(B_1) = (0, 0, 0)$ and $maj(B_2) = (0, 1, 0)$, we have that $B_2 \succ^3 B_1$ and that $B_1$, unlike $B_2$, comprises a truthful strategy by 3.}*

The fact that truthful voting is not always a dominant strategy for JA games with payoff 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 from 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 6, a straightforward adaptation of the proof of Proposition 3 proves the following:

**Proposition 7.** Every consistent $A^\pi$-game for maj has an IC-consistent NE that is truthful and goal-efficient.

Observe that aggregation games can be modelled as a special subclass of $A^\pi$-games where each $\pi_i$ is a constant function for each $i \in \mathcal{N}$. Therefore the negative results expressed in Proposition 4 still hold in this case. In particular, we can still find undesirable equilibria (e.g., untruthful, goal-inefficient and IC-inconsistent).
4.2 Equilibria in games with non-uniform payoff

Despite the previous observations, we can use non-uniform payoffs to rule out goal-inefficient NE, as shown next:

**Proposition 8.** For every consistent uniform $A^\pi$-game $\langle A, \{\pi_i\}_{i \in N}\rangle$ for $\text{maj}$ there exist payoff functions $\{\pi'_i\}_{i \in N}$ such that $\pi'$ is a redistribution of payoffs given by $\pi$, i.e., $\sum_{i \in N} \pi'_i(B) = \sum_{i \in N} \pi_i(B)$ for every profile $B$, and $\langle A, \{\pi'_i\}_{i \in N}\rangle$ has no goal-inefficient NE.

**Proof.** Let $B^*$ be a ballot such that $B^* \models \bigwedge_i \gamma_i$. We now construct a redistribution of payoffs in which player 1 gives all other players 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 $\langle A, \{\pi_i\}_{i \in N}\rangle$. The desired payoff function is constructed as follows. For all $j \neq 1$ define $\pi'_j(B) = \pi_j(B) + M$ for all profiles $B$ with $B_j = B^*$, and $\pi'_j(B) = \pi_j(B)$ otherwise. Let finally $\pi'_1(B) = \pi_1(B) - (\sum_{k \neq 1} \pi'_k(B) - \sum_{k \neq 1} \pi_k(B))$. Observe that the construction of $\pi'$ ensures that $\sum_{i \in N} \pi'_i(B) = \sum_{i \in N} \pi_i(B)$, for every profile $B$. Now let $B$ be a goal-inefficient NE of the new game. Take an arbitrary player $j$ such that $B_j \neq B^*$ (such a player always exists due to $\text{maj}$ and the fact that $B$ is goal inefficient). By construction of $\pi'_j$, player $j$ has an incentive to deviate to $B^*$, hence $B$ cannot be a NE.

The proposition says that a payoff function exists which eliminates a goal-inefficient NE in an original, uniform, $A^\pi$-game, while keeping the sum of players’ payoffs constant. The new payoff function can be thought of as an offer of payoff that a player has made to the others, incentivising them to deviate to a more favourable outcome.

## 5 Endogenous Aggregation Games

In this section we introduce the full-fledged model of pre-vote negotiations. In a nutshell, the game has two phases (in line with Jackson and Wilkie, [3]):

- A pre-vote phase, where, starting from a uniform $A^\pi$-game, players make simultaneous transfers of payoff to their fellow players;

- A vote phase, where players play the original $A^\pi$-game, updated with transfers.

We call these extended voting games endogenous aggregation games (henceforth $A^T$-games).

**Definition 9 ($A^T$-games).** An endogenous aggregation game is defined as a tuple $\langle A, \{\pi_i\}_{i \in N}, \{T_i\}_{i \in N}\rangle$ where $\langle A, \{\pi_i\}_{i \in N}\rangle$ is a uniform $A^\pi$ game, and each $T_i$ is the set of all transfer functions $\tau_i : \text{Mod}(IC)^N \times N \to \mathbb{R}_+$. 
Expressions such as $\tau_i(B, j)$ encode the amount of payoff that a player $i$ gives to player $j$ should a certain profile of votes $B$ be played. We call $\tau \in \prod_i T_i$ a transfer profile, denoting $\tau^0$ the void transfer where at every profile every player gives 0 to the others. So by $\tau(A^\tau) = \langle A, \{\pi'_i\}_{i \in \mathcal{N}} \rangle$ we denote the aggregation game with payoff obtained from $A^\tau$ where $\pi'_i$ is updated according to the transfer profile $\tau$. $A^\tau$-games are analysed as extensive form games with perfect information and simultaneous moves. The solution concept of choice will therefore be subgame perfect equilibrium. Strategies of player $i$ consist of sequential choices of a transfer $\tau_i$ and an admissible ballot $B_i$. Preference relations are naturally defined between tuples of the form $(\tau, B)$: for two profiles of ballots $B, B'$, the expression $(\tau, B) \succeq_\pi (\tau', B')$ denotes the fact that player $i$ prefers $B$ after $\tau$ has been played in the pre-vote phase, to $B'$ after $\tau'$ has been played in the pre-vote phase.

It is important to notice that transfers do not preserve the uniformity of payoffs. There is actually always a transfer that turns a uniform $A^\pi$-game into a non uniform one. In particular, while we know that a uniform consistent game for majority always has a Nash equilibrium, this may no longer be the case after a transfer. To overcome this limitation, rather than allowing mixed strategies, we simply assume that any deviation to a game $\tau(A)$ with no pure strategy Nash equilibrium is never profitable for a player. This might seem a rather demanding limitation, as there might be infinitely many transfer functions that lead to games with no pure strategy Nash equilibria. However, we have already seen that uniform consistent $A^\pi$-games for majority always have pure strategy equilibria (cf. Proposition 7) and the questions we are interested in answering concern the sustainability of those equilibria or their replacement with new ones after the pre-vote phase.

Given a $A^\tau$-game $\langle A, \{\pi_i\}_{i \in \mathcal{N}}, \{T_i\}_{i \in \mathcal{N}} \rangle$ we call a Nash equilibrium $B$ of $\langle A, \{\pi_i\}_{i \in \mathcal{N}} \rangle$ a surviving Nash equilibrium if there exists a transfer function $\tau$ and a subgame perfect equilibrium of $A^\tau$ where $(\tau, B)$ is played on the equilibrium path. In our case, subgame perfect equilibria are constructed selecting (i) a pure strategy Nash equilibrium after each transfer profile, whenever it exists, and (ii) a transfer profile, such that no profitable deviation exist for a player by changing her individual transfer function.

Surviving equilibria identify those electoral outcomes that can intuitively be rationally sustained by an appropriate pre-vote negotiation.

### 5.1 Equilibria in Endogenous JA Games

Pre-vote negotiations have desirable consequences for the players, in particular goal-efficient NE of a $A^\pi$-game are all surviving in the induced endogenous aggregation game.

**Proposition 9.** Let $A^\tau = \langle A, \{\pi_i\}_{i \in \mathcal{N}}, \{T_i\}_{i \in \mathcal{N}} \rangle$ be an endogenous aggregation game with more than two players. Every goal-efficient NE of $\langle A, \{\pi_i\}_{i \in \mathcal{N}} \rangle$ is a surviving NE.
Proof (sketch). Let $B$ be a goal-efficient NE of $\langle A, \{\pi_i\}_{i \in N} \rangle$. We want to find a transfer function $\tau^*$ such that $(\tau^*, B)$ is a subgame perfect equilibrium of $A^T$. Let $M - 1$ be the maximal payoff difference, as defined in the proof of Proposition 8. For all $i$, let $\tau^*_i(B', j) = 2M$ if $B'_i \neq B_i$, and $\tau^*_i(B', j) = 0$ otherwise. In words, each player $i$ is committing to play the ballot $B_i$ by offering the others $2M$ in case she deviates. The fact that $(\tau^*, B)$ is a subgame perfect equilibrium in the two-phase game follows from the fact that $B$ is goal-efficient (i.e., all deviations are payoff based) and from the standard argument for pre-play negotiations with more than two players ([3], Theorem 4).

Notice that Proposition 9 carries through independently of the voting procedure adopted in the underlying aggregation game. A stronger result holds for $maj$:

**Proposition 10.** Let $A^T = \langle A, \{\pi_i\}_{i \in N}, \{T_i\}_{i \in N} \rangle$ be an endogenous aggregation game for $maj$ such that $A$ is consistent. Every surviving NE of $A^T$ is goal-efficient.

Proof. Let $B^*$ be a NE that is goal-inefficient, i.e., such that $maj(B^*) \not\models \gamma_i$ for some individual $i$, and assume for the sake of contradiction that $B^*$ is a surviving equilibrium. Therefore there exists a subgame perfect equilibrium of $A^T$ such that $(\tau^*, B^*)$ is played on the equilibrium path. We now construct a profitable deviation from $\tau^*$, leading to contradiction. By consistency of $A$ there exists a ballot $B'$ such that $B' \models \bigwedge_{j \in N} \gamma_j$, hence in particular $B' \models \gamma_i$. Let now $i$ deviate to a transfer profile $\tau' = (\tau'_i, \tau_{-i}^*)$ such that she offers an amount $M -$ as defined in the proof of Proposition 8—to all other players if they vote for ballot $B'$. This transfer will make $B'$ a dominant strategy for all players. Consider now the updated game $\tau'(A^*)$. Because of $maj$, $B'$ will be the outcome of every NE, and one NE, e.g., the unanimous vote for $B'$, always exists. Hence $\tau'$ is a profitable deviation for $i$ since $\gamma_i$ is satisfied in every such NE.

Summing up, pre-vote negotiations are a powerful tool players have to overcome the inefficiencies of aggregation procedures. In particular, when players’ goals 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 want to point out that players’ equilibrium strategies in the two-phase game remain individually rational strategies and the game remains non-cooperative throughout, even when equilibrium strategy end up sustaining efficiency.

With pre-vote negotiations we can also look at classical paradoxes such as the discursive dilemma under a new light.

**Example 3.** Consider again Table 4 and let the parties’ goals be: $\gamma'_A = W, \gamma'_B = \neg F, \gamma'_C = N$. Let $\pi$ be a uniform payoff function such that the discursive dilemma is a NE. This equilibrium is not surviving because party $C$ could transfer enough utility to party $B$ for it to vote for $N$ and be better off in the resulting game.
With respect to this, it is also worthwhile to note the following direct consequence of Proposition 7 and Theorem 9:

**Corollary 11.** Every consistent $A^*-game for maj has an IC-consistent NE that is truthful, goal-efficient and survives in $A^T = (A^*, \{T_i\}_{i \in N})$.

In consistent aggregation games equilibria that give rise to a voting paradox may not survive (Example 3), whereas equilibria avoiding such paradoxes are always sustained by a pre-vote negotiation phase (Corollary 11).

### 6 Conclusions

We have shown the effect of having a pre-vote phase before an aggregation game, where voters might hold a principled opinion about a subset of the issues at stake while willing to strike deals on the remaining ones and can influence one another before casting their ballots in order to obtain an individually more favourable electoral outcome.

We have seen how undesirable equilibria, i.e., where players’ goals are not satisfied, can be eliminated if players’ goals are consistent by appropriate transfers, while desirable ones, i.e., where players’ goals are satisfied, can be rationally sustained.

Our model can be applied to the study of a variety of opinion change phenomena such as: deliberation, negotiation, power and authority, bribery, etc., where voters invest resources (e.g., persuasiveness, money, time, personal credibility, etc.) to influence one another.

### References

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

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

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

[4] Kenneth Arrow. *Social Choice and Individual Values*. Yale University Press, 1951.

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

[6] Amrita Dhillon and Ben Lockwood. When are plurality rule voting games dominance-solvable? *Games and Economic Behavior*, 46(1):55 – 75, 2004.
[7] 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.

[8] 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.

[9] Michael Wooldridge, Ulle Endriss, Sarit Kraus, and Jérôme Lang. Incentive engineering for boolean games. *Artificial Intelligence*, 195:418–439, 2013.

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