

Pre-Vote Negotiations in Binary Aggregation

Umberto Grandi and Davide Grossi and Paolo Turrini

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 [5, 10]. Indeed, in strategic interaction pre-play negotiations are known to be effective in overcoming inefficient allocations caused by players' individual rationality [9]. 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 [9].

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 [1], 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 on three of them:

Judgment Aggregation. Judgment aggregation is the theory of collective decision-making on logically interconnected issues [4, 7]. We take here the perspective of a society 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 integrity constraints [6]. With the exception of manipulability considerations [12, 3], the fully fledged 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 [2] and Messner and Polborn [11].

Boolean games. We model decision strategies in judgment aggregation as Boolean Games [8, 15], 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 utility-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 [14], which studies boolean games together with a pre-play phase. In the spirit of Jackson and Wilkie [9] players are entitled to sacrifice a part of the utility 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 collective 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 utilities, analyzing the properties of the equilibria (Section 3). Second, we add an explicit utility 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* [6]. 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.

$$\text{IC} = (W \wedge \neg F) \rightarrow P$$

| | 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

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

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

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

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 $\text{IC} = (W \rightarrow P)$, making ballot $(P \mapsto 1, W \mapsto 0)$ —in words, voting positively for W and negatively for P —an inadmissible ballot.

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

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

Examples of aggregation functions include the *issue-by-issue majority rule* (*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 [6].

Given a BA structure \mathcal{S} , an aggregation procedure F is said to be *collectively rational* with respect to \mathcal{S} if the outcome $F(\mathbf{B})$ is an admissible ballot for every admissible profile \mathbf{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 $\mathcal{A} = \langle \mathcal{N}, \mathcal{I}, \text{IC}, F, \{\gamma_i\}_{i \in \mathcal{N}} \rangle$ such that:*

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

All individuals share the same set of admissible strategies, namely the set of admissible ballots $\text{Mod}(\text{IC})$. A strategy profile is therefore a profile of (admissible) binary ballots, and will be denoted with \mathbf{B} . Preferences in aggregation games are *only* formulated in terms of goal satisfaction. We will refer to an aggregation game $\mathcal{A} = \langle \mathcal{N}, \mathcal{I}, \text{IC}, F, \{\gamma_i\}_{i \in \mathcal{N}} \rangle$ where $F = \text{maj}$ as an *aggregation game for maj* (or for the majority rule).

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

It clearly follows that $\succeq_i^{\mathcal{A}}$ is a total preorder. Following standard terminology from voting theory we call a strategy (or ballot) $B \in \text{Mod}(\text{IC})$ *truthful for agent i* if it is in accordance with i 's preferences, that is, if it satisfies i 's goal γ_i . We call a game *consistent* if the conjunction of all individual goals $\bigwedge_i \gamma_i$ is consistent with IC.

Definition 5. *Given an aggregation game \mathcal{A} , we call a strategy profile $\mathbf{B} = (B_1, \dots, B_n)$:*

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

3.1 Equilibria in Aggregation Games

In this section we explore the existence and the properties of pure Nash equilibria (NE) in aggregation games and their properties, paying special attention to NE that are truthful, goal-efficient and IC-consistent. We rely on standard game-theoretic terminology from [13]. 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 . We want to show that for each strategy profile \mathbf{B}_{-i} we have that $(\mathbf{B}_{-i}, B_i^*) \succeq_i^{\mathcal{A}} (\mathbf{B}_{-i}, B_i')$. We have four cases. If both $\text{maj}(\mathbf{B}_{-i}, B_i^*)$ and $\text{maj}(\mathbf{B}_{-i}, B_i')$ satisfy γ_i or both do not satisfy it, then B_i^* weakly dominates B_i' . If $\text{maj}(\mathbf{B}_{-i}, B_i^*)$ satisfies γ_i and $\text{maj}(\mathbf{B}_{-i}, B_i')$ does not, then B_i^* strictly dominates B_i' . Finally, if $\text{maj}(\mathbf{B}_{-i}, B_i')$ satisfies γ_i but $\text{maj}(\mathbf{B}_{-i}, B_i^*)$ does not, then the majority for γ_i is larger in (\mathbf{B}_{-i}, B_i') than in (\mathbf{B}_{-i}, B_i^*) , against the assumption that B_i^* is truthful. Hence B_i^* is weakly dominant. \square

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

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 $\mathbf{B}^* = (B^*, \dots, B^*)$, i.e., \mathbf{B}^* is a unanimous profile where each individual votes for B^* . Since the profile is unanimous, $maj(\mathbf{B}^*) = B^*$. Clearly $maj(\mathbf{B}^*)$ satisfies $\bigwedge_i \gamma_i$ and IC, and each individual votes truthfully. Hence \mathbf{B}^* is a NE by Proposition 1. \square

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.*

| | p_1 | p_2 | p_3 |
|----------|-------|-------|-------|
| Voter 1 | 1 | 0 | 0 |
| Voter 2 | 0 | 1 | 0 |
| Voter 3 | 0 | 0 | 1 |
| Majority | 0 | 0 | 0 |

Table 2: IC-inefficient equilibria

Proof. For the first statement, let \mathcal{A} be an aggregation game for maj such that $\gamma_i = p_i$, let $IC = \top$ and let \mathbf{B}^* be the profile illustrated in Table 2. We can observe that \mathbf{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 \mathcal{A}' with the same individual goals but $IC' = p_1 \vee p_2 \vee p_3$. Then \mathbf{B}^* is a truthful NE which is both goal-inefficient and IC-inconsistent. \square

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:*

$$\begin{aligned}\gamma_1 &= W \wedge \neg F \wedge P \\ \gamma_2 &= W \wedge F \wedge \neg P \\ \gamma_3 &= \neg W \wedge \neg F \wedge \neg P\end{aligned}$$

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.

4 Aggregation Games with Utility

This section refines the model of aggregation games, introducing an explicit utility 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 6 (A^π games). *An aggregation game with utility is a tuple $\langle \mathcal{A}, \{\pi_i\}_{i \in \mathcal{N}} \rangle$ where \mathcal{A} is an aggregation game and $\pi_i : \text{Mod}(\text{IC})^{\mathcal{N}} \rightarrow \mathbb{R}$ is a utility function assigning to each admissible profile a real number denoting the utility of player i in that profile.*

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 utility function. This is technically called a *quasi-dichotomous* preference relation [15]:

Definition 7 (Preferences in A^π games). *Let $\langle \mathcal{A}, \{\pi_i\}_{i \in \mathcal{N}} \rangle$ be an A^π -game, \mathbf{B}, \mathbf{B}' be two ballot profiles and $i \in \mathcal{N}$ a player. The preference relation \succeq_i^π for each $i \in \mathcal{N}$ is such that $\mathbf{B} \succeq_i^\pi \mathbf{B}'$ iff:*

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

In other words, a profile \mathbf{B} is preferred by player i to \mathbf{B}' if either $F(\mathbf{B})$ satisfies i 's goal and \mathbf{B}' doesn't or, if both satisfy i 's goal or neither do, $F(\mathbf{B})$ yields to i a better utility than $F(\mathbf{B}')$.

A natural class of A^π -games is that of games where the individual utility only depends on the outcome of the collective decision, that is, for all $i \in \mathcal{N}$, $\pi_i(\mathbf{B}) = \pi_i(\mathbf{B}')$ whenever $F(\mathbf{B}) = F(\mathbf{B}')$. We call these games *uniform*.

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

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

4.1 Equilibria in Games with Uniform Utility

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

Proposition 5. *There exist uniform A^π -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 utility function as follows, let $\pi_i(B) = 1$ for $i = 3$ and $B = (0, 1, 0)$, and 0 otherwise. Take the following profiles: $\mathbf{B}_1 = ((0, 1, 0), (0, 0, 0), (0, 0, 1))$ and $\mathbf{B}_2 = ((0, 1, 0), (0, 0, 0), (0, 1, 0))$. Since $\text{maj}(\mathbf{B}_1) = (0, 0, 0)$ and $\text{maj}(\mathbf{B}_2) = (0, 1, 0)$, we have that $\mathbf{B}_2 \succ_3^\pi \mathbf{B}_1$ and that \mathbf{B}_1 , unlike \mathbf{B}_2 , comprises a truthful strategy by 3. \square

The fact that truthful voting is not always a dominant strategy for JA games with utility might seem counterintuitive, especially when the utility 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 utility.

Despite the negative result in Proposition 5, a straightforward adaptation of the proof of Proposition 3 to the case of dichotomous preferences with uniform utilities proves the following:

Proposition 6. *Every consistent uniform A^π -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^π -games where each π_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 (goal-inefficient and IC-inconsistent).

4.2 Equilibria in Games with Non-Uniform Utility

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

Proposition 7. *For every consistent uniform A^π -game $\langle \mathcal{A}, \{\pi_i\}_{i \in \mathcal{N}} \rangle$ for maj there exist utility functions $\{\pi'_i\}_{i \in \mathcal{N}}$ such that π' is a redistribution of utilities given by π , i.e., $\sum_{i \in \mathcal{N}} \pi'_i(\mathbf{B}) = \sum_{i \in \mathcal{N}} \pi_i(\mathbf{B})$ for every profile \mathbf{B} , and $\langle \mathcal{A}, \{\pi'_i\}_{i \in \mathcal{N}} \rangle$ 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 utilities 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 utility difference that some player can obtain between two outcomes in the game $\langle \mathcal{A}, \{\pi_i\}_{i \in \mathcal{N}} \rangle$. The desired utility function is constructed as follows. For all $j \neq 1$ define $\pi'_j(\mathbf{B}) = \pi_j(\mathbf{B}) + M$ for all profiles \mathbf{B} with $B_j = B^*$, and $\pi'_j(\mathbf{B}) = \pi_j(\mathbf{B})$ otherwise. Let finally $\pi'_1(\mathbf{B}) = \pi_1(\mathbf{B}) - (\sum_{k \neq 1} \pi'_k(\mathbf{B}) - \sum_{k \neq 1} \pi_k(\mathbf{B}))$. Observe that the construction of π' ensures that $\sum_{i \in \mathcal{N}} \pi'_i(\mathbf{B}) = \sum_{i \in \mathcal{N}} \pi_i(\mathbf{B})$, for every profile \mathbf{B} . Now let $\bar{\mathbf{B}}$ be a goal-inefficient NE of the new game. Take an arbitrary player j such that $\bar{B}_j \neq B^*$ (such a player always exists due to maj and the fact that $\bar{\mathbf{B}}$ is goal inefficient). By construction of π'_j , player j has an incentive to deviate to B^* , hence $\bar{\mathbf{B}}$ cannot be a NE. \square

The proposition says that a utility function exists which eliminates a goal-inefficient NE in an original, uniform, A^π -game, while keeping the sum of players' utilities constant. The new utility function can be thought of as an *offer of utility* 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 [9]):

- A *pre-vote phase*, where, starting from a uniform A^π -game, players make simultaneous transfers of utility to their fellow players;
- A *vote phase*, where players play the original A^π -game, updated with transfers.

We call these extended voting games *endogenous aggregation games* (henceforth \mathcal{A}^T -games).

Definition 8 (\mathcal{A}^T -games). *An endogenous aggregation game is defined as a tuple $\langle \mathcal{A}, \{\pi_i\}_{i \in \mathcal{N}}, \{T_i\}_{i \in \mathcal{N}} \rangle$ where $\langle \mathcal{A}, \{\pi_i\}_{i \in \mathcal{N}} \rangle$ is a uniform A^π -game, and each T_i is the set of all transfer functions $\tau_i : \text{Mod}(\text{IC})^{\mathcal{N}} \times \mathcal{N} \rightarrow \mathbb{R}_+$*

Expressions such as $\tau_i(\mathbf{B}, j)$ encode the amount of utility that a player i gives to player j should a certain profile of votes \mathbf{B} be played. We call $\tau \in \prod_i T_i$ a *transfer profile*, denoting τ^0 the *void* transfer where at every profile every player gives 0 to the others. So by $\tau(A^\pi) = \langle \mathcal{A}, \{\pi'_i\}_{i \in \mathcal{N}} \rangle$ we denote the aggregation game with utility obtained from A^π where π'_i is updated according to the transfer profile τ [9]. \mathcal{A}^T -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 τ_i and an admissible ballot B_i . Preference relations are naturally defined between tuples of the form (τ, \mathbf{B}) : for two profiles of ballots \mathbf{B}, \mathbf{B}' , the expression $(\tau, \mathbf{B}) \succeq_i^\pi (\tau', \mathbf{B}')$ denotes the fact that player i prefers \mathbf{B} after τ has been played in the pre-vote phase, to \mathbf{B}' after τ' has been played in the pre-vote phase.

It is important to notice that transfers do not preserve the uniformity of utilities. There is actually always a transfer that turns a uniform A^π -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 $\tau(\mathcal{A})$ 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^π -games for majority always have pure strategy equilibria (cf. Proposition 6) 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 \mathcal{A}^T -game $\langle \mathcal{A}, \{\pi_i\}_{i \in \mathcal{N}}, \{T_i\}_{i \in \mathcal{N}} \rangle$ we call a Nash equilibrium \mathbf{B} of $\langle \mathcal{A}, \{\pi_i\}_{i \in \mathcal{N}} \rangle$ a *surviving Nash equilibrium* if there exists a transfer function τ and a subgame perfect equilibrium (cf. [13]) of \mathcal{A}^T where (τ, \mathbf{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 Aggregation Games

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

Proposition 8. *Let $\mathcal{A}^T = \langle \mathcal{A}, \{\pi_i\}_{i \in \mathcal{N}}, \{T_i\}_{i \in \mathcal{N}} \rangle$ be an endogenous aggregation game with more than two players. Every goal-efficient NE of $\langle \mathcal{A}, \{\pi_i\}_{i \in \mathcal{N}} \rangle$ is a surviving NE.*

Proof (sketch). Let \mathbf{B} be a goal-efficient NE of $\langle \mathcal{A}, \{\pi_i\}_{i \in \mathcal{N}} \rangle$. We want to find a transfer function τ^* such that (τ^*, \mathbf{B}) is a subgame perfect equilibrium of \mathcal{A}^T . Let $M - 1$ be the maximal utility difference, as defined in the proof of Proposition 7. For all i , let $\tau_i^*(\mathbf{B}', j) = 2M$ if $B'_i \neq B_i$, and $\tau_i^*(\mathbf{B}', j) = 0$ 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 (τ^*, \mathbf{B}) is a subgame perfect equilibrium in the two-phase game follows from the fact that \mathbf{B} is goal-efficient (i.e., all deviations are utility based) and from the standard argument for pre-play negotiations with more than two players ([9], Theorem 4). \square

Notice that Proposition 8 carries through independently of the voting procedure adopted in the underlying aggregation game. A stronger result holds for *maj*:

Proposition 9. *Let $\mathcal{A}^T = \langle \mathcal{A}, \{\pi_i\}_{i \in \mathcal{N}}, \{T_i\}_{i \in \mathcal{N}} \rangle$ be an endogenous aggregation game for *maj* such that \mathcal{A} is consistent. Every surviving NE of \mathcal{A}^T is goal-efficient.*

Proof. Let \mathbf{B}^* be a NE that is goal-inefficient, i.e., such that $\text{maj}(\mathbf{B}^*) \not\models \gamma_i$ for some individual i , and assume for the sake of contradiction that \mathbf{B}^* is a surviving equilibrium. Therefore there exists a subgame perfect equilibrium of \mathcal{A}^T such that (τ^*, \mathbf{B}^*) is played on the equilibrium path. We now construct a profitable deviation from τ^* , leading to contradiction. By consistency of \mathcal{A} there exists a ballot B' such that $B' \models \bigwedge_{j \in \mathcal{N}} \gamma_j$, 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 7—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 $\tau'(\mathcal{A}^\pi)$. Because of *maj*, B' will be the outcome of every NE, and one NE, e.g., the unanimous vote for B' , always exists. Hence τ' is a profitable deviation for i since γ_i is satisfied in every such NE. \square

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 in a new light.

Example 3. Consider again Table 1 and let the parties' goals be: $\gamma'_A = W, \gamma'_B = \neg F, \gamma'_C = N$. Let π be a uniform utility 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 6 and Theorem 8:

Corollary 10. Every consistent A^π -game for *maj* has an IC-consistent NE that is truthful, goal-efficient and survives in $\mathcal{A}^T = \langle A^\pi, \{T_i\}_{i \in \mathcal{N}} \rangle$.

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 10).

6 Conclusions

We have shown the effect of having a pre-vote phase before an aggregation game. Voters may 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. We believe our model can be applied to the study of a variety of opinion change phenomena, albeit at a high level of abstraction, 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.

The model lends itself to pursue several lines of research. First of all, it will be important to study results concerning the survivability of equilibria in games that are not consistent.

Second, integrity constraints have not been a central focus of the current paper but lie at the heart of aggregation problems. In future work we want to investigate further how pre-vote negotiation can effectively support collectively rational aggregation. Finally, the model can be enriched along several dimensions in order to make it more realistic. In this respect the addition of budgets for players' transfers seems to be a natural extension of the model.

Acknowledgments

We thank the anonymous reviewers for their helpful suggestions. Paolo Turrini acknowledges the support of the IEF Marie Curie fellowship Norms in Action: Designing and Comparing Regulatory Mechanisms for Multi-Agent Systems (FP7- PEOPLE-2012-IEF, 327424 NINA).

References

- [1] Kenneth Arrow. *Social Choice and Individual Values*. Yale University Press, 1951.
- [2] Amrita Dhillon and Ben Lockwood. When are plurality rule voting games dominance-solvable? *Games and Economic Behavior*, 46(1):55 – 75, 2004.
- [3] F. Dietrich and C. List. Strategy-proof judgment aggregation. *Economics and Philosophy*, 23:269–300, 2007.
- [4] F. Dietrich and C. List. Propositionwise judgment aggregation: The general case. *Social Choice and Welfare*, 40:1067–1095, 2013.
- [5] John Dryzek and Christian List. Social choice theory and deliberative democracy: A reconciliation. *British Journal of Political Science*, 33:1–28, 2003.
- [6] Umberto Grandi and Ulle Endriss. Lifting integrity constraints in binary aggregation. *Artificial Intelligence*, 199-200:45–66, 2013.
- [7] D. Grossi and G. Pigozzi. *Judgment Aggregation: A Primer*. Morgan and Claypool Publishers, 2014.
- [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] Matthew O. Jackson and Simon Wilkie. Endogenous games and mechanisms: Side payments among players. *Review of Economic Studies*, 72(2):543–566, 2005.
- [10] Christian List. Group communication and the transformation of judgments: An impossibility result. *The Journal of Political Philosophy*, 19(1):1–27, 2011.
- [11] 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.
- [12] K. Nehring and C. Puppe. Strategy-proof social choice on single-peaked domains: Possibility, impossibility and the space between. Working Paper, University of California at Davis, 2002.
- [13] M. J. Osborne and A. Rubinstein. *A Course in Game Theory*. MIT Press, 1994.
- [14] Paolo Turrini. Endogenous boolean games. In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI-2013)*, 2013.

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

Umberto Grandi
Department of Mathematics
University of Padova
Italy
Email: umberto.uni@gmail.com

Davide Grossi
Department of Computer Science
University of Liverpool
United Kingdom
Email: d.grossi@liverpool.ac.uk

Paolo Turrini
Department of Computing
Imperial College London
United Kingdom
Email: p.turrini@imperial.ac.uk