Mediators and Truthful Voting

Bezalel Peleg[*] Ariel D. Procaccia[†]

Abstract

The Gibbard-Satterthwaite Theorem asserts the impossibility of designing a non-dictatorial social choice function (SCF) under which agents can never benefit by lying, assuming complete information. We show that in voting games of complete information where a mediator is on hand, the implications of this troubling impossibility result can be alleviated. Indeed, we characterize families of SCFs where truth-telling is a *mediated equilibrium* under any constellation of preferences; under such an equilibrium, the outcome is enforced by punishing deviating agents with the help of a mediator. In particular, we observe that the family of feasible elimination procedures has the foregoing property.

1 Introduction

We consider a setting which consists of a set of *agents* and a set of *alternatives*; the agents have linear preferences over the alternatives. A *social choice function* (SCF) receives the preferences of the agents as input, and outputs the winning alternative. Informally, the celebrated Gibbard-Satterthwaite (G-S) Theorem [4, 14] asserts that, if it is the case that agents never have an incentive to lie, then the SCF must be dictatorial, that is, one agent dictates the outcome of the election regardless of the preferences of others.

More precisely, let us denote the set of agents by $N = \{1, \ldots, n\}$, the preferences of agent i by R^i, and a *preference profile* for all the agents by $R^N = \langle R^1, \ldots, R^n \rangle$. One can represent an SCF F as a *game form* Γ^F. The strategies of agents are simply their possible rankings R^i of alternatives; for each combination of strategies (rankings), the outcome is the one specified by the SCF (see Table 1 for an example). We can now define the *game associated with* Γ^F *and* R^N as a concept akin to games in normal form: the outcome is determined by Γ^F, but instead of utilities, each agent i has a preference relation R^i over the possible outcomes.

The G-S Theorem is usually stated in terms of truth-telling as a dominant strategy. However, in our setting it is easily verified that the theorem can be reformulated as follows: for every non-dictatorial SCF F (with range of size at least 3), there is a preference profile R^N that is not a Nash equilibrium point of the game associated with Γ^F and R^N. A Nash equilibrium is a strategy profile such that no agent can gain by unilaterally deviating. We will find this somewhat unorthodox, but nevertheless accurate, statement of the theorem more helpful in the sequel.

It is important to note that the G-S Theorem deals with games of complete information. In other words, the result holds only if the manipulator has complete knowledge of the preferences of the other agents. We shall deal with the complete information setting throughout the paper.

[*]Institute of Mathematics and Center for the Study of Rationality, The Hebrew University of Jerusalem. email: pelegba@math.huji.ac.il

[†]School of Engineering and Applied Sciences, Harvard University, email: arielpro@seas.harvard.edu. The work was done while the author was at the Hebrew University of Jerusalem.

	xR^2y	yR^2x
xR^1y	x	x
yR^1x	x	y

xR^3y

	xR^2y	yR^2x
xR^1y	x	y
yR^1x	y	y

yR^3x

Table 1: The Plurality SCF as a game form. In Plurality, an alternative wins if it is ranked first by more agents than any other alternative. In this example we have 3 agents and 2 alternatives (x and y). Each agent i has two strategies: xR^iy or yR^ix. Agent 1 chooses top or bottom row, agent 2 chooses left or right column, and agent 3 chooses left or right matrix.

Our approach. In this paper we propose that the implications of the G-S Theorem can be alleviated to some degree by employing mediators. Broadly speaking, instead of seeking to implement the truthful outcome as a Nash equilibrium, we design SCFs such that truth-telling is always a *mediated equilibrium* (to be defined later). This stands in contrast to the impossibility of designing nondictatorial SCFs where truth-telling is always a Nash equilibrium (the G-S Theorem). That is, we achieve a possibility result by tweaking the solution concept.

In the next few paragraphs, we shall attempt to outline the main ideas of our work via several examples. As a first example, consider the Borda SCF with 3 agents and 5 alternatives. Under this rule, each agent gives 4 points to its top ranked alternatives, 3 points to its second choice, and so on. The preference profile R^N is defined below on the left hand side:

R^1	R^2	R^3
a	b	d
b	a	c
c	c	b
d	d	a
e	e	e

Q^1	R^2	R^3
a	b	d
e	a	c
c	c	b
d	d	a
b	e	e

Q^1	Q^2	Q^3
a	e	e
e	d	d
c	c	c
d	b	b
b	a	a

Alternative b is the winner under R^N, with 9 points. Given its information about the preferences of agents 2 and 3, agent 1 readily recognizes that it is better off expressing the preferences Q^1, thus obtaining the preference profile (Q^1, R^2, R^3) given above in the middle. In this case, a is the winner with 8 points while b has only 6 points, and the former alternative is preferred by agent 1 to b.

Now, suppose agents 2 and 3, observing agent 1's deceit, had the ability to punish agent 1 by changing their votes to Q^2 and Q^3, respectively, as above on the right hand side. Under this ballot, e is the winner—the worst alternative as far as agent 1 is concerned. So, if agent 1 knew that if it plays any preferences other than R^1, agent 2 and 3 would surely "press the red button"[1] by playing Q^2 and Q^3, agent 1 would have an incentive to tell the truth.

This is exactly the idea underlying Aumann's *acceptable* strategy profiles [1, 2]. Aumann's concept is more ambitious, as it deals with deviations of coalitions.

> Here it is not sufficient that the players of B be able to increase their payoffs by changing their strategies while the players of $N \setminus B$ maintain their strategies as they are. For

[1] We use the term "red button" here as the outcome e is also bad for agents 2 and 3.

the players of $N \setminus B$ will not in general maintain their strategies fixed in the face of a change on the part of B, but will also change their strategies in order to meet the new conditions; thus B's glory would be short-lived indeed [1, Section 4].

So, informally, a strategy profile is acceptable if a deviating coalition can always be punished by a deviation of the remaining agents. Aumann's original reasoning stemmed from infinite-horizon *repeated games* of complete information, where deviators can be punished by the complement in a future time. Since we deal with *one-shot games*, the power of agents to punish deviating coalitions must be interpreted in a different way.

One possible interpretation is to assume that the agents employ a mediator. Cooperative agents give the mediator the right to play for them. If some of the agents deviate, the mediator plays for the remaining agents a predefined strategy profile that depends on the observed identities (or even actions) of the deviating agents, thereby punishing the deviators. Since the mediator is preconfigured, the "threat" of punishment is sufficient to deter potential deviators. We note that there is also an explicit interpretation in terms of threats, which is due to Pattanaik [7, 8]. He considers a voting setting where a coalition threatens to deviate, and the complement responds with a counter-threat. Pattanaik argues that in some situations (for example, when the voting is public) agents would be aware of the formation of a deviating coalition, which would allow the complement to react. Nevertheless, we feel that the interpretation in terms of mediators is intuitive and adopt it in the sequel.

Monderer and Tennenholz [5] argue that the concept of mediator is central and natural in computational settings, and accordingly develop a solution concept for games in normal form which they call *mediated equilibria*. The formal distinction between mediated equilibria and acceptable strategies is somewhat subtle, and does not survive the transition to game forms, where only pure strategies are considered. Nevertheless, the reasoning underlying the work of Monderer and Tennenholz also holds in our setting. We therefore define a mediated equilibrium in game forms to be a strategy profile where every deviating coalition can be punished by its complement. Mediated equilibria are also featured in our subsequent, very recently published work [11]; please see Section 5 for more details.

Let us recapitulate. Given a setting (e.g., a mediator) where deviators can be punished, it seems easy to discourage manipulations by a single agent. Indeed, in any reasonable SCF a vast majority of agents can force any outcome, and in particular punish the manipulator by switching to an outcome that is bad for it. But what about deviating coalitions?

Let us consider another example: 5 agents, 4 alternatives, and the Plurality rule (each agent gives a point to its top choice). A mediator can be configured as follows: if a single agent decides not to give the mediator the right of play, all other agents cast their vote in favor of the worst alternative of the deviator. Observe the preference profile on the left hand side:

R^1	R^2	R^3	R^4	R^5
a	a	b	c	d
b	b	c	b	c
c	c	d	d	b
d	d	a	a	a

R^1	R^2	Q^3	Q^4	Q^5
a	a	b	b	b
b	b	c	c	c
c	c	d	d	d
d	d	a	a	a

Alternative a wins this election, with 2 points. Agents 3, 4 and 5 would be better served by agreeing on some other alternative, such as b, so they deviate to the profile on the right hand side. Now, can agents 1 and 2 exact their vengeance? The answer is no. In Plurality, a minority cannot overturn a unanimous majority decision. More generally, to the best of our knowledge all the common SCFs are susceptible to manipulation by coalitions, even in our setting where punishments are possible. Our goal in this paper is therefore to design SCFs where truth-telling is a mediated equilibrium.

Overview of results. We say that a social choice function F is *truthful in mediated equilibrium* if every preference profile R^N is a mediated equilibrium of the game associated with Γ^F and R^N. In other words, in the game (of complete information) associated with Γ^F and R^N, the ability to punish deviators (possibly enforced by a mediator) incentivizes the agents to play the truthful preference profile R^N.

Our main theorems in fact provide a recipe that can be used to produce SCFs that are truthful in mediated equilibrium. The SCFs that we construct are based on *effectivity functions*—mappings from coalitions of agents to families of subsets of alternatives. We show that, given an effectivity function that satisfies certain reasonable conditions, it is possible to find a corresponding SCF where truth-telling is a mediated equilibrium, even when the mediator cannot observe the actions of deviating coalitions. We also consider the setting where the mediator is better informed about the strategies of the deviators, and show that it is possible to obtain a different family of SCFs that are truthful in (a different type of) mediated equilibrium. Interestingly, in this case we obtain rules where, in addition to ranking the alternatives, the agents also have to choose an integer each; these integers are used to determine the exact social choice among a set of desirable alternatives.

The SCFs that are truthful in mediated equilibrium can also be designed to be socially sensible, and in particular to satisfy specific desiderata like Pareto-optimality and monotonicity. As an example of the applicability of our results, we show that the intuitive family of feasible elimination procedures is truthful in mediated equilibrium.

Structure of the paper. The paper proceeds as follows. In Section 2 we give some background about game forms and effectivity functions. In Section 3, we adapt the notions of mediated equilibria to our social choice setting. In Section 4, we construct families of SCFs that are truthful in mediated equilibrium. We discuss our results in Section 5.

2 Preliminaries

In this section we define and explain relevant notions from social choice theory. For a more detailed exposition of these subjects, the reader may consult [9].

2.1 Game Forms

For a set K, we denote by $\mathcal{P}(K)$ the powerset of K (the set of all subsets of K), and by $\mathcal{P}_0(K)$ the set of all nonempty subsets of K.

We consider a set $N = \{1, \dots, n\}$ of *agents*, and a set $A = \{x_1, \dots, x_m\}$ of *alternatives*. The agents have linear preferences over alternatives; formally, for every agent $i \in N$ we have a binary relation R^i over A which satisfies antisymmetry, transitivity and totality. We will also assume that R^i satisfies reflexivity, i.e. xR^ix for all $x \in A$. Any such relation R^i induces a strict preference relation P^i, given by: xP^iy if and only if xR^iy and $x \neq y$. Denote by $\mathcal{L} = \mathcal{L}(A)$ the set of all linear preferences over A—we may write $R^i \in \mathcal{L}$. $R^N = \langle R^1, \ldots, R^n \rangle \in \mathcal{L}^N$ represents a *preferences profile* of all the agents, and R^S represents the preferences of coalition $S \in \mathcal{P}_0(N)$. We sometimes write xR^Sy if xR^iy for all $i \in S$.

A *game form* (GF) is an $(n+1)$-tuple $\Gamma = \langle \Sigma^1, \ldots, \Sigma^n; \pi \rangle$, where Σ^i, $i = 1, \ldots, n$, is a nonempty finite set, and $\pi : \Sigma^N \to A$. Σ^i is called the set of *strategies* of agent i, and π is the *outcome function.* We will assume that π is onto A.

A game form is a "pregame", in the sense that there are no incentives associated with it. Incentives come into play when a game form is coupled with a preference profile. Formally, the *game associated with* Γ *and* R^N is the n-person game in normal form

$$g(\Gamma, R^N) = \langle \Sigma^1, \ldots, \Sigma^n; \pi; R^1, \ldots, R^n \rangle.$$

In the game $g(\Gamma, R^N)$, combinations of agents' strategies determine an outcome; instead of having utilities over outcomes, agents are assumed to possess a more abstract preference relation.

Once a game form is coupled with incentives, one can consider some of the solution concepts that are prevalent in games in normal form. A Nash equilibrium is a strategy profile where no agent can benefit by unilaterally deviating. Formally, $\sigma^N \in \Sigma^N$ is a *Nash equilibrium point* of $g(\Gamma, R^N)$ if for every $i \in N$ and every $\tau^i \in \Sigma^i$, $\pi(\sigma^N) R^i \pi(\tau^i, \sigma^{N \setminus \{i\}})$.

Let us now briefly present a game-theoretic perspective on social choice. A *social choice function (SCF)* is a function $F : \mathcal{L}^N \to A$, and a *social choice correspondence* is a function $H : \mathcal{L}^N \to \mathcal{P}_0(A)$. Notice that an SCF F is equivalent to the GF $\Gamma^F = \langle \mathcal{L}, \ldots, \mathcal{L}; F \rangle$; indeed, in this game form the agents' strategies are orderings of alternatives, and the outcome is determined by F. For an example, see Table 1.

We now wish to translate the standard formulation of the Gibbard-Satterthwaite theorem into the language of game forms. Informally, the Gibbard-Satterthwaite Theorem states that under any non-dictatorial SCF with range of size at least three, truth-telling is *not* a dominant strategy, namely there are situations in which agents may benefit by lying. In our setting, it is easy to verify that this statement is equivalent to the following formulation using Nash equilibria.

Theorem 2.1 (Gibbard-Satterthwaite [4, 14])**.** *Let* $F : \mathcal{L}^N \to A$ *be an SCF onto* A, $|A| \geq 3$. *If for every* $R^N \in \mathcal{L}^N$, R^N *is a Nash equilibrium point of* $g(\Gamma^F, R^N)$, *then* F *is dictatorial (there exists a agent* i *such that* $F(R^N)$ *is always the alternative ranked first in* R^i*).*

We insist on this nonstandard formulation since it highlights the contrast with our results. Indeed, we show in the sequel that there exist nondictatorial SCFs such that for every $R^N \in \mathcal{L}^N$, R^N is a *mediated equilibrium* of $g(\Gamma^F, R^N)$.

2.2 Effectivity Functions

An *effectivity function* (EF) is a function $E : \mathcal{P}_0(N) \to \mathcal{P}(\mathcal{P}_0(A))$ such that for every $S \in \mathcal{P}_0(N)$, $A \in E(S)$, and for every $B \in \mathcal{P}_0(A)$, $B \in E(N)$. So, effectivity functions map coalitions of players to sets of subsets of alternatives. In an abstract sense, such functions represents the power distribution among individuals in a society. If a subset $B \in \mathcal{P}_0(A)$ satisfies $B \in E(S)$, where E is an effectivity function, we say that S is *effective* for B. Conceptually, this means that the players in B can force the outcome to be one of the alternatives in B.

Different notions of what it means to "force the outcome" induce different effectivity functions. In this paper, we will deal with only three effectivity functions. α-effectiveness implies that the agents in S can coordinate their strategies so that, no matter what the other agents do, the outcome will be in B. Formally, S *is α-effective for* B if there exists $\sigma^S \in \Sigma^S$ such that for all $\tau^{N\setminus S} \in \Sigma^{N\setminus S}$, $\pi(\sigma^S, \tau^{N\setminus S}) \in B$. The *$\alpha$-effectivity function* associated with a game form Γ is given by

$$E_\alpha^\Gamma(S) = \{B \in \mathcal{P}_0(A) : \ S \text{ is } \alpha\text{-effective for } B\}.$$

On the other hand, if S is β-effective for B, the agents in S can counter any action profile of $N \setminus S$ with actions of their own such that the outcome is in B. More precisely, S *is β-effective for* B if for every $\tau^{N\setminus S} \in \Sigma^{N\setminus S}$ there exists $\sigma^S \in \Sigma^S$ such that $\pi(\sigma^S, \tau^{N\setminus S}) \in B$. The *$\beta$-effectivity function associated with* Γ is given by

$$E_\beta^\Gamma(S) = \{B \in \mathcal{P}_0(A) : \ S \text{ is } \beta\text{-effective for } B\}.$$

Clearly α-effectivity is stronger than β-effectivity

The α-effectivity function of an SCC $H : L^N \to \mathcal{P}_0(A)$, denoted E_α^H, is defined similarly: S is α-effective for B if there exists $R^S \in \mathcal{L}^S$ such that for all $Q^{N\setminus S} \in \mathcal{L}^{N\setminus S}$, $H(R^S, Q^{N\setminus S}) \subseteq B$. The β-effectivity function of H, denoted E_β^H, is again defined in the obvious way. Next we define the last effectivity that we shall require. Let $H : L^N \to \mathcal{P}_0(A)$ be an SCC, $S \in \mathcal{P}_0(N), B \in \mathcal{P}_0(A)$. S is *winning* for B if and only if for all $R^N \in L^N$,

$$[\forall x \in B, \forall y \notin B, \ xR^S y] \Rightarrow H(R^N) \subseteq B \quad .$$

The *first EF* associated with H is the function $E_*^H : \mathcal{P}_0(N) \to \mathcal{P}(\mathcal{P}_0(A))$ defined by

$$E_*^H(S) = \{B \in \mathcal{P}_0(A) : \ S \text{ is winning for } B\}.$$

The core of an EF E is a set outcomes such that no coalition has both the power and the will to force a different outcome. Formally, we say that $B \in \mathcal{P}_0(A \setminus \{x\})$ *dominates* $x \in A$ *via* $S \in \mathcal{P}_0(N)$ if $B \in E(S)$ and $BP^S x$. B *dominates* x if there exists $S \in \mathcal{P}_0(N)$ such that B dominates x via S. The *core* of E is the set of undominated alternatives in A, and is denoted by $C(E; R^N)$. If B dominates x via S, x is (in a sense) unstable, as the agents in S can force the outcome to be in B, and prefer any alternative in B to x. So, in this sense, the core of an effectivity function is the set of stable alternatives. We say that an effectivity function is *stable* if its core is always nonempty, namely for all $R^N \in \mathcal{L}^N$, $C(E; R^N) \neq \emptyset$.

We now list some additional properties of EFs that we will require in the sequel. An EF E is *monotonic with respect to the agents* if and only if for every $S \in \mathcal{P}_0(N)$ and $B \in E(S)$, if $S \subseteq T$ then $B \in E(T)$. E is *monotonic with respect to the alternatives* if and only if for every $S \in \mathcal{P}_0(N)$ and $B \in E(S)$, if $B \subseteq B^*$ then $B^* \in E(S)$. E is *monotonic* if and only if it is monotonic with respect to both agents and alternatives.

An EF E is *superadditive* if and only if for every $S_i \in \mathcal{P}_0(N), B_i \in E(S_i)$, $i = 1, 2$, if $S_1 \cap S_2 = \emptyset$ then $B_1 \cap B_2 \in E(S_1 \cup S_2)$. In words, if S_1 is effective for B_1 and S_2 is effective for B_2, then their union is effective for the intersection of B_1 and B_2.

Finally, E is *maximal* if and only if for every $S \in \mathcal{P}_0(N)$ and $B \in \mathcal{P}_0(A)$, if $B \notin E(S)$ then $A \setminus B \in E(N \setminus S)$, that is, if S is not effective for B then its complement is effective for the complement of B.

3 A Characterization of Mediated Equilibria in Game Forms

Our contribution begins by defining analogs of the concept of mediated equilibria for game forms. In this section we assume the reader has already gained some intuitions regarding the ability to punish deviators, and its interpretation in one-shot games using mediators, by reading Section 1, and therefore we will be able to quickly reach the formal definitions.

A notion of mediators that is stronger than the one described in Monderer and Tennenholz [5] was investigated by Rozenfeld and Tennenholz [13]. In this work, the mediator may have access to different levels of information. One setting is when mediators can observe the selected actions of all the agents, not merely the ones that empower them. The motivation behind this definition comes from the world of communication networks: a router obtains information about messages passing through it, including their routes. So, in network games, such *routing mediators* [13] appear natural.[2] Hence we may consider two types of mediators, based on the level of information available to them.

- *Simple mediators*: These mediators are not aware of the strategies of agents that do not use their services, only their identity.
- *Informed mediators*: Such mediators (e.g., routers) are aware of the joint strategy of the deviating coalition $N \setminus S$.

This, in turn, motivates the consideration of two different methods of punishment. The simple setting is the exact equivalent of the notion of acceptable equilibria in game forms: the punishing coalition must have a joint strategy that discourages its complement from deviating, regardless of the joint strategy of the complement. In the informed setting, the punishing coalition has complete knowledge of the strategies of the complement (e.g., via an informed mediator), and can base the punishment on this knowledge. The latter setting is very similar to the one considered by Pattanaik [7, 8]. An important remark is that by "punishing a deviating coalition" we mean that at least one of the members of the coalition does not gain from the deviation, and therefore has no incentive to participate.

Definition 3.1. Let $\Gamma = \langle \Sigma^1, \ldots, \Sigma^n; \pi \rangle$ be a game form, $R^N \in \mathcal{L}^N$.

1. $\sigma^N \in \Sigma^N$ is a *simple mediated* equilibrium of the game $g(\Gamma, R^N)$ if and only if for every $S \in \mathcal{P}_0(N)$ there exists $\tau^S \in \Sigma^S$ such that for every $\tau^{N \setminus S} \in \Sigma^{N \setminus S}$ there exists $i \in N \setminus S$ such that $\pi(\sigma^N) R^i \pi(\tau^N)$.
2. $\sigma^N \in \Sigma^N$ is an *informed mediated* equilibrium of the game $g(\Gamma, R^N)$ if and only if for every $S \in \mathcal{P}_0(N)$ and $\tau^{N \setminus S} \in \Sigma^{N \setminus S}$ there exist $\tau^S \in \Sigma^S$ and $i \in N \setminus S$ such that $\pi(\sigma^N) R^i \pi(\tau^N)$.

Note that the question of whether a strategy profile is a mediated equilibrium depends only on the outcome of the strategy profile, not on the profile itself. In other words, if some strategy profile is a simple (resp., informed) mediated equilibrium then any other strategy profile with the same outcome is also a simple (resp., informed) mediated equilibrium.

The following lemma characterizes mediated equilibria via the cores of effectivity functions. It essentially says that being in the core of the α-effectivity function is the same as being the outcome

[2] Some recent applications of routing mediators can be found in [12].

of an informed mediated equilibrium, while being in the core of the β-effectivity function is the same as being the outcome of a simple mediated equilibrium. This brings to the forefront an interesting analogy between the two levels of information manifested in both effectivity functions and mediators.

Lemma 3.2. *Let $\Gamma = \langle \Sigma^1, \ldots, \Sigma^n; \pi \rangle$ be a game form such that π is onto A, $R^N \in \mathcal{L}^N$, $x \in A$.*

1. *$x \in C(E^{\Gamma}_{\alpha}; R^N)$ if and only if every $\sigma^N \in \Sigma^N$ such that $\pi(\sigma^N) = x$ is an informed mediated equilibrium of $g(\Gamma, R^N)$.*
2. *$x \in C(E^{\Gamma}_{\beta}; R^N)$ if and only if every $\sigma^N \in \Sigma^N$ such that $\pi(\sigma^N) = x$ is a simple mediated equilibrium of $g(\Gamma, R^N)$.*

Proof. For part 1, assume first that $x \in C(E^{\Gamma}_{\alpha}; R^N)$; assume there exists $\sigma^N \in \Sigma^N$ such that $\pi(\sigma^N) = x$, but σ^N is not an informed mediated equilibrium of $g(\Gamma, R^N)$. Then there exist $S \in \mathcal{P}_0(N), \tau^{N \setminus S} \in \Sigma^{N \setminus S}$ such that for all $\tau^S \in \Sigma^S$ and $i \in N \setminus S$, $\pi(\tau^N) P^i \pi(\sigma^N) = x$. Let

$$B = \pi(\Sigma^S, \tau^{N \setminus S}) = \{y \in A : \exists \tau^S \in \Sigma^S \ s.t. \ \pi(\tau^S, \tau^{N \setminus S}) = y\}.$$

We have that B dominates x via $N \setminus S$, in contradiction to our assumption that $x \in C(E^{\Gamma}_{\alpha}; R^N)$.

In the other direction, let $\sigma^N \in \mathcal{L}^N$ such that $\pi(\sigma^N) = x$ (there exists such a strategy profile as π is onto), and assume σ^N is an informed mediated equilibrium, but $x \notin C(E^{\Gamma}_{\alpha}; R^N)$. Then there exist $B \in \mathcal{P}_0(A)$ and $S \in \mathcal{P}_0(N)$ such that B dominates x via $N \setminus S$. In other words, there is $\tau^{N \setminus S} \in \Sigma^{N \setminus S}$ such that for all $\tau^S \in \Sigma^S$ it holds that $\pi(\tau) \in B$, and in addition the agents in $N \setminus S$ prefer B to $x = \pi(\sigma^N)$. This is a contradiction to the assumption that σ^N is an informed mediated equilibrium.

The proof of part 2 is an easy reformulation of the proof of part 1. □

4 Construction of SCFs that are Truthful in Mediated Equilibrium

In this section we design SCFs where it is possible to enforce truth-telling given the ability to punish deviators via a mediator (and given complete information, as in the rest of the paper). Of course, we also have to take into account the information the mediator can access.

Definition 4.1. $F : \mathcal{L}^N \to A$ is *truthful in simple (informed) mediated equilibrium* if for all $R^N \in \mathcal{L}^N$, R^N is a simple (informed) mediated equilibrium in $g(\Gamma^F, R^N)$.

If we show the existence of such social choice functions, then we take an important step towards alleviating the implications of the Gibbard-Satterthwaite Impossibility result [4, 14]. Indeed, given an SCF that is truthful in mediated equilibrium, no *coalition* is motivated to deviate from R^N in the game $g(\Gamma^F, R^N)$, as the mediator is guaranteed to adjust the strategies of the complement in a way which worsens the outcome for at least one member of the deviating coalition. Since in the game $g(\Gamma^F, R^N)$ in fact R^N is the truthful preference profile, it is essentially possible to enforce truth-telling in this game.

4.1 Truthfulness in Simple Mediated Equilibrium

We first wish to establish the existence of SCFs that are truthful in simple mediated equilibrium. We prove the following theorem:

Theorem 4.2. *Let E be a stable and maximal effectivity function, and let $F : \mathcal{L}^N \to A$ such that $F(R^N) \in C(E; R^N)$ for all $R^N \in \mathcal{L}^N$. Then $E^F_\beta = E$.*

In the formulation of the theorem, E^F_β is a shorthand for $E^{\Gamma^F}_\beta$. What is the implication of this theorem? Consider an SCF F that is a choice from the core of a stable and maximal effectivity function, as required in the theorem. Since $E^F_\beta = E$, we have that for any $R^N \in \mathcal{L}^N$, $C(E; R^N) = C(E^F_\beta; R^N)$. So F is in fact a choice from $C(E^F_\beta; R^N)$, but we know from Lemma 3.2 that any preference profile that yields an outcome in $C(E^F_\beta; R^N)$ is a simple mediated equilibrium in $g(\Gamma^F, R^N)$. Hence, we obtain the following immediate corollary to Theorem 4.2:

Corollary 4.3. *Let E be a stable and maximal effectivity function, and let $F : \mathcal{L}^N \to A$ such that $F(R^N) \in C(E; R^N)$ for all $R^N \in \mathcal{L}^N$. Then F is truthful in simple mediated equilibrium.*

Before we prove Theorem 4.2, we require some simple lemmas. The next lemma, however, is slightly more general than what we need at the moment; we will make use of the general formulation in the sequel. It considers a generalization of the concept of SCFs: game forms where the set of strategies of agent i includes a ranking of the alternatives, and a choice of action from a more restricted set Σ^i_1. In particular, if Σ^i_1 is a singleton, we effectively obtain a game form that is equivalent to an SCF. We note that, given two EFs E and E', $E \subseteq E'$ means that for all $S \in \mathcal{P}_0(N)$ and $B \in \mathcal{P}_0(A)$, if $B \in E(S)$ then $B \in E'(S)$.

Lemma 4.4. *Let E be a stable effectivity function, and let Γ be a game form such that for all $i \in N$, $\Sigma^i = \mathcal{L} \times \Sigma^i_1$, and $\pi(R^1, \sigma^1_1, \ldots, R^n, \sigma^n_1) \in C(E; R^N)$. Then $E \subseteq E^\Gamma_\alpha$.*

Proof. Let $S \in \mathcal{P}_0(N)$ and $B \in \mathcal{P}_0(A)$ such that $B \in E(S)$. We have to show that $B \in E^\Gamma_\alpha(S)$. Indeed, let $R^S \in \mathcal{L}^S$ such that B is at the top of R^i for all $i \in S$, and fix some $\sigma^S_1 \in \Sigma^S_1$. Now, let $Q^{N\setminus S} \in \mathcal{L}^{N\setminus S}, \sigma^{N\setminus S}_1 \in \Sigma^{N\setminus S}_1$, and $x = \pi(R^S, \sigma^S_1, Q^{N\setminus S}, \sigma^{N\setminus S}_1)$. It must be the case that $x \in B$; otherwise, BP^ix for all $i \in S$ and $B \in E(S)$, or in other words B dominates x via S, in contradiction to the construction of π as a choice from $C(E; R^S, Q^{N\setminus S})$. □

Lemma 4.5. *Let E be a maximal effectivity function, and let E' be a superadditive effectivity function. If $E \subseteq E'$ then $E = E'$.*

Proof. Assume that there exist $S \in \mathcal{P}_0(N)$ and $B \in \mathcal{P}_0(A)$ such that $B \in E'(S) \setminus E(S)$. As E is maximal, it follows that $A \setminus B \in E(N \setminus S)$, and therefore $A \setminus B \in E'(N \setminus S)$. By the super-additivity of E' we obtain that

$$\emptyset = B \cap (A \setminus B) \in E'(S \cup (N \setminus S)) = E'(N).$$

This is a contradiction to the definition of E' as an effectivity function. □

Lemma 4.6. *[9, Lemma 5.1.17] Let Γ be a game form. Then $E^\Gamma_\alpha = E^\Gamma_\beta$ if and only if E^Γ_α is maximal.*

We are now ready to prove Theorem 4.2.

Proof of Theorem 4.2. By Lemma 4.4, $E \subseteq E^F_\alpha$. By the assumption that E is maximal and the trivial observation that E^F_α is superadditive, it follows from Lemma 4.5 that $E = E^F_\alpha$. From Lemma 4.6 we have that $E^F_\alpha = E^F_\beta$, and thus $E^F_\beta = E$. □

So, Theorem 4.2 basically gives a formula to derive SCFs that are truthful in simple mediated equilibrium from *any* effectivity function that is stable and maximal. The construction has two degrees of freedom: the choice of E (such that E is stable and maximal), and the exact (deterministic) choice from the core of E. In the next few paragraphs, we discuss a natural family of SCFs that satisfies the conditions of the theorem.

Let $\beta : A \to \{1, \dots, n\}$ such that $\sum_{j=1}^m \beta(x_j) = n + 1$, and let $R^N \in \mathcal{L}^N$. A *feasible elimination procedure* with respect to R^N and $\beta(\cdot)$ is a sequence $(x_{j_1}, S_1; \dots; x_{j_{m-1}}, S_{m-1}; x_{j_m})$ (where $S_1, \dots, S_{m-1}$ are coalitions) that satisfies the following:

1. For all $s \neq t$, $S_s \cap S_t = \emptyset$.
2. For all $t = 1, \dots, m-1$, $|S_t| = \beta(x_{j_t})$.
3. For all $t = 1, \dots, m-1$, $\forall i \in S_t$, x_{j_t} is the worst alternative with respect to R^i restricted to $\{x_{j_t}, \dots, x_{j_m}\}$.

In a feasible elimination procedure, each coalition S_t intuitively vetoes its worst alternative, out of the alternatives which have not been vetoed so far. The alternative that survives the process, i.e., has not been vetoed by any coalition, wins the election.

Example 4.7. Let $N = \{1, 2\}$, $A = \{a, b, c\}$, and $\beta(a) = \beta(b) = \beta(c) = 1$. Define F as follows: agent 1 vetoes its worst alternatives, then agent 2 vetoes the alternative it favors less out of the two remaining alternatives. F always yields a feasible elimination procedure. Indeed, consider the preference profile R^N where:

R^1	R^2
b	a
a	b
c	c

It holds that $F(R^N) = a$, as agent 1 vetoes c, then agent 2 vetoes b. This is a feasible elimination procedure given by $(c, \{1\}; b, \{2\}; a)$.

Peleg [9] proves that a feasible elimination procedure is a choice from the core of the effectivity function E_f given by:

$$B \in E_f(S) \Leftrightarrow |S| \geq \beta(A \setminus B).$$

Conversely, every choice from the core of the effectivity function E_f is obtained by a feasible elimination procedure. Peleg also proves that E_f is stable and maximal. Therefore, by Theorem 4.2, we have:

Corollary 4.8. *The family of feasible elimination procedures is truthful in simple mediated equilibrium.*

In addition, Peleg [9] gives different sufficient conditions for effectivity functions that are stable and maximal. For example, an effectivity function is *additive* if there exist positive probability measures p on A and q on N such that for every $B \in \mathcal{P}_0(A)$ and $S \in \mathcal{P}_0(N)$,

$$B \in E(S) \Leftrightarrow q(S) > 1 - p(B).$$

Peleg proves that additive effectivity functions are stable, and shows that an additive effectivity function is maximal if and only if

$$[B \in \mathcal{P}_0(A), B \neq A, S \in \mathcal{P}_0(N), S \neq N] \Rightarrow p(B) \neq q(S).$$

Such conditions can be used to directly construct SCFs which are truthful in simple mediated equilibrium, but these rules may not be easy to grasp intuitively (unlike feasible elimination procedures).

The SCFs derived from Theorem 4.2 can be made to satisfy some of the prominent desiderata in social choice. First, any SCF derived according to the theorem is Pareto-optimal, i.e., if all the agents rank $x \in A$ above $y \in A$ then y cannot be elected. Indeed, in this case x dominates y via N, so $y \notin C(E; R^N)$—but $F(R^N) \in C(E; R^N)$ for all $R^N \in \mathcal{L}^N$.

Second, as noted above the theorem gives some leeway in the exact choice from the core; there is a choice that is guaranteed to be *monotonic*, i.e., if $F(R^N) = x$ and Q^N is obtained from R^N by pushing x upwards in the agents' preferences, then $F(Q^N) = x$. This choice is obtained by fixing some $R_0 \in \mathcal{L}$, and always choosing the alternative in the core which is ranked highest by R_0. It follows from Peleg [9, Lemma 2.4.8] and the properties of the core that this choice is monotonic.

4.2 Truthfulness in Informed Mediated Equilibrium

In order to obtain different families of SCFs that are truthful in mediated equilibrium, we are going to eschew the assumption made in Theorem 4.2 that E is maximal. However, we will have to compensate by weakening the theorem in several other aspects. First, we will obtain SCFs that are truthful in *informed* mediated equilibrium (instead of simple).

Second, instead of an SCF in the accepted sense, we shall construct a generalized SCF, where the agents express their preferences by ranking the alternatives (as usual), but in addition each agent also chooses an integer in $\{1, ..., m\}$. The outcome of the election is determined by both the rankings and the integers. However, as we shall see, the rankings effectively designate the set of acceptable alternatives (the core of an effectivity function), and the integers only serve to choose among the acceptable alternatives.

Third, we will make additional assumptions regarding E, specifically that E is monotonic and supperadditive (very weak assumptions) and that E satisfies $E_\alpha^{C(E;\cdot)} \subseteq E_*^{C(E;\cdot)}$, that is, the α-effectivity function of the core of E is contained in the first effectivity function of the core of E (we discuss this assumption in the sequel). We wish to prove:

Theorem 4.9. *Let E be a stable, monotonic, and superadditive EF that satisfies $E_\alpha^{C(E;\cdot)} \subseteq E_*^{C(E;\cdot)}$. Then there exists a game form*

$$\Gamma = \langle \mathcal{L} \times \{1, \ldots, m\}, \ldots, \mathcal{L} \times \{1, \ldots, m\}; \pi \rangle$$

such that

1. *For all* $(R^N, r^N) \in \Sigma^N$*,* $\pi(R^N, r^N) \in C(E; R^N)$*.*

2. $E^\Gamma_\alpha = E$*.*

We shall require the following lemma.

Lemma 4.10. *[9, Lemma 6.1.21] Let E be a stable and monotonic EF. If* $H(R^N) = C(E; R^N)$ *for every* $R^N \in \mathcal{L}^N$*, then* $E^H_* = E$*.*

Proof of Theorem 4.9. We start with the construction of Γ. In order to define π, let $R_0 \in \mathcal{L}^N$. Given a strategy profile $\sigma^N = \langle (R^1, r^1), \cdots, (R^n, r^n) \rangle$, let $r \equiv r^1 + \cdots + r^n (\text{mod } m)$. We let

$$j_0 = \min \ \operatorname{argmin}_{j \in \{1,\dots,m\}:\ x_j \in C(E;R^N)} |j - (r+1)|.$$

In other words, we choose the index j_0 which is closest to $r+1$ among the indices of alternatives in the core of E. If there are two such indices, the smaller is chosen. Now, the outcome is $\pi(\sigma^N) = x_{j_0}$.

By Lemma 4.4, we have that

$$E \subseteq E^\Gamma_\alpha. \tag{1}$$

We shall prove inclusion in the opposite direction. We first claim that

$$E^\Gamma_\alpha \subseteq E^{C(E;\cdot)}_\alpha. \tag{2}$$

Indeed, let $S_0 \in \mathcal{P}_0(N)$ and $B_0 \in \mathcal{P}_0(A)$ such that $B_0 \in E^\Gamma_\alpha(S_0)$. More explicitly, this means that

$$\exists R^{S_0}_0 \in \mathcal{L}^{S_0}, r^{S_0}_0 \ \textit{s.t.}\ \forall Q^{N\setminus S_0} \in \mathcal{L}^{N\setminus S_0}, q^{N\setminus S_0}, \pi(R^{S_0}_0, r^{S_0}_0, Q^{N\setminus S_0}, q^{N\setminus S_0}) \in B_0. \tag{3}$$

Note that if $S_0 = N$ then (2) follows trivially, hence can assume that $S_0 \neq N$. We will show that for all $Q^{N\setminus S_0} \in \mathcal{L}^{N\setminus S_0}$, $C(E; R^{S_0}_0, Q^{N\setminus S_0}) \subseteq B_0$. Assume that there exist $Q^{N\setminus S_0} \in \mathcal{L}^{N\setminus S_0}$ and $x_{j_0} \in C(E; R^{S_0}_0, Q^{N\setminus S_0})$ such that $x_{j_0} \notin B_0$. Since $N \setminus S_0 \neq \emptyset$, there exists $q^{N\setminus S_0}$ such that

$$\sum_{i\in S_0} r^i_0 + \sum_{i \in N\setminus S_0} q^i \equiv j_0 - 1 (\text{mod } m).$$

By the definition of π, we have that

$$\pi(R^{S_0}_0, r^{S_0}_0, Q^{N\setminus S_0}, q^{N\setminus S_0}) = x_{j_0} \notin B_0.$$

This is a contradiction to (3).

We can now complete the proof by relying on a known equality and our assumption. From (2) we have that $E^\Gamma_\alpha \subseteq E^{C(E;\cdot)}_\alpha$. By our assumption, $E^{C(E;\cdot)}_\alpha \subseteq E^{C(E;\cdot)}_*$. By Lemma 4.10, under our assumptions it holds that $E^{C(E;\cdot)}_* = E$. Hence $E^\Gamma_\alpha \subseteq E$, and using (1) it follows that $E^\Gamma_\alpha = E$, as claimed. □

So, the set of strategies Σ^i of agent i consists of $R^i \in \mathcal{L}$—a ranking of alternatives—and some number $r^i \in \{1, \dots, m\}$. It follows from the proof that in Theorem 4.9 we do not have any leeway regarding which choice from the core to embrace. Rather, the agents determine this by submitting numbers r^i. The alternative in the core with index closest to $\sum_{i=1}^n r^i (\text{mod } m) + 1$ is the outcome. This way, any agent, given the others' votes, can single-handedly determine which of the (desirable) alternatives in the core is chosen; this is crucial for our proof.

Remark 4.11. Peleg [10] shows that, given an effectivity function E which is superadditive and monotonic, it is possible to obtain a game form Γ where $E = E^{\Gamma}_{\alpha}$. However, Theorem 4.9 deals with a game form that is very close to an SCF, while the game form introduced by Peleg is much more general and therefore not similar to an SCF.

Like before, Lemma 3.2 gives us the immediate corollary (stated somewhat informally):

Corollary 4.12. *Let E be a stable, monotonic, and superadditive effectivity function that satisfies $E_{\alpha}^{C(E;\cdot)} \subseteq E_{*}^{C(E;\cdot)}$. Then there exists a generalized social choice function Γ, where agents choose rankings of alternatives $R^i \in \mathcal{L}$ and an integer $r^i \in \{1, \ldots, m\}$, and the outcome is a choice from $C(E; R^N)$, such that (R^N, r^N) is always an informed mediated equilibrium of $g(\Gamma, R^N)$.*

Consider, for example, the effectivity function defined by $E(N) = \mathcal{P}_0(A)$, and $E(S) = \{A\}$ for all $S \subsetneq N$. It is easy to see that this function is monotonic and superadditive. In addition, E is stable; indeed, $C(E; R^N)$ is exactly the set of Pareto-optimal alternatives in R^N, i.e.,

$$C(E; R^N) = \{x \in A : \nexists y \in A \ s.t. \ \forall i \in N, yP^i x\}.$$

In particular, this set is never empty since it includes any alternative which some agent ranks first. This also makes it straightforward to verify that $E_{\alpha}^{C(E;\cdot)} \subseteq E_{*}^{C(E;\cdot)}$. So, if we in fact use the theorem to derive a (generalized) SCF from the foregoing effectivity function E, we can obtain a socially sensible SCF that is truthful in informed mediated equilibrium. This SCF is a choice from exactly the set of Pareto-optimal alternatives, and it is anonymous (indifferent to the identity of the agents). Moreover, notice that E in this example is not maximal, so Theorem 4.2 does not apply.

More generally, it is possible to show that the game form Γ of Theorem 4.9 always satisfies an appropriate notion of Pareto-optimality. Indeed, if for all $i \in N$ it holds that $xP^i y$, then $y \notin C(E; R^N)$, as x dominates y via N. The desired property follows since π is a choice from $C(E; R^N)$.

Finally, a short discussion of the properties of the effectivity function assumed in Theorem 4.9 is in order. Requiring that E be stable, monotonic, and superadditive is not very restrictive. We do, however, have to justify the assumption that $E_{\alpha}^{C(E;\cdot)} \subseteq E_{*}^{C(E;\cdot)}$.

A simple game is a pair $G = (N, W)$ where N is a set of agents, and W is the set of *winning coalitions*, which satisfies:

$$S \in W \wedge S \subseteq T \Rightarrow T \in W.$$

Given a simple game G where $N \in W$, the *standard* effectivity function associated with G is given by $E(S) = \mathcal{P}_0(A)$ if $S \in W$, and $E(S) = \{A\}$ otherwise [9, Definition 6.2.1]. It can be verified that any effectivity function associated with a simple game satisfies $E_{\alpha}^{C(E;\cdot)} \subseteq E_{*}^{C(E;\cdot)}$. Since simple games are ubiquitous in real life settings (e.g., committees and parliaments), it can be argued that this assumption is natural as well.

Remark 4.13. One may wonder whether it is always the case that $E_{\alpha}^{C(E;\cdot)} \subseteq E_{*}^{C(E;\cdot)}$ for any effectivity function E that is stable, monotonic, and superadditive. We can answer this question in the negative using the following counterexample.

Let $N = \{1, \ldots, 4\}$ and $A = \{x_1, x_2, x_3, y\}$. Define an effectivity function E as follows. For all $i \in N$, $E(\{i\}) = \{A\}$. The subset $\{1, 2\}$ is effective for $\{x_2, y\}$ and every superset; the subset $\{2, 3\}$ is effective for $\{x_3, y\}$ and every superset; the subset $\{1, 3\}$ is effective for $\{x_1, y\}$ and every superset. If $|S| = 2, 4 \in S$, then $E(S) = \{A\}$. The subset $\{1, 2, 3\}$ is effective for $\{x_1, y\}$, $\{x_2, y\}$, $\{x_3, y\}$, and every superset. If $|S| = 3$, $4 \in S$, then $E(S) = E(S \setminus \{4\})$. Finally, by definition $E(N) = \mathcal{P}_0(A)$.

E is monotonic and superadditive by construction. We claim that E is stable. Indeed, let R^N such that there is no $x \in A$ such that $xP^N y$ (i.e., y is not Pareto-dominated); then $y \in C(E; R^N)$. Otherwise, let x such that $xP^N y$ and x is not Pareto-dominated, then $x \in C(E; R^N)$.

We want to show that $E_\alpha^{C(E;\cdot)} \nsubseteq E_*^{C(E;\cdot)}$. Since $E_*^{C(E,\cdot)} = E$ by Lemma 4.10, it suffices to show that there exist $S_0 \in \mathcal{P}_0(N), B_0 \in \mathcal{P}_0(A)$ such that $B_0 \in E_\alpha^{C(E;\cdot)}(S_0)$ but $B_0 \notin E(S_0)$. Let $B_0 = \{y\}$, $S_0 = \{1, 2, 3\}$. It holds that $B_0 \notin E(S_0)$.

To show that $B_0 \in E_\alpha^{C(E,\cdot)}(S_0)$, we have to show that there exists $R^{S_0} \in \mathcal{L}^{S_0}$ such that for every $R^{N \setminus S_0} \in \mathcal{L}^{N \setminus S_0}$ and for every $x \notin B_0$ there exist $S \in \mathcal{P}_0(N)$ and $B \in \mathcal{P}_0(A)$ such that $BR^S x$ and $B \in E(S)$. Define R^{S_0} by:

R^1	R^2	R^3
y	y	y
x_2	x_3	x_1
x_1	x_2	x_3
x_3	x_1	x_2

Then $\{y, x_2\}$ dominates x_1 via $\{1, 2\}$, $\{y, x_3\}$ dominates x_2 via $\{2, 3\}$, and $\{y, x_1\}$ dominates x_3 via $\{1, 3\}$.

5 Discussion

In this section we discuss some prominent issues.

Relation to Implementation Theory and subsequent work. The implementation problem can be described as follows: given an SCC H, design a game form such that for every preference profile R^N, the game's outcomes in equilibrium under R^N are exactly the set $H(R^N)$. Our work differs in two main respects from straightforward implementation by mediated equilibrium. First, we ask that that truth-telling be a mediated equilibrium, but there may be other mediated equilibria with different outcomes (this is sometimes called *partial implementation*). Second, and crucially, we demand that the implementing game form be an SCF, rather than some general game form with possibly nonintuitive strategy sets, as is often the case in implementation results. Our notion of implementation is referred to as *truthful implementation* by Osborne and Rubinstein [6, pp. 179–180].

In more recent work, which builds on a draft of this paper, we deal with implementation by mediated equilibrium in the classic sense [11]. In particular, we provide characterizations that imply that prominent SCCs that are not implementable by strong equilibrium are implementable by mediated equilibrium.

The complete information assumption. In order to punish deviators, the mediator must be configured in a way that requires complete information about the truthful preferences of the agents. The complete information assumption might seem irrelevant in many voting situations. A critic might therefore suggest that our model is flawed. Our answer to this claim is twofold.

First, complete information is typical in different voting settings, such as decision problems for small committees with a discussion period prior to voting. In this discussion period, the agents can configure the mediator so that agents do not deviate when the actual election takes place. Moreover, in the other extreme, in large anonymous voting games a great deal of information is typically available to the voters. For example, in mass elections only the distribution of the votes matters, and it is often known from polls.

Second, and more importantly, our goal in this paper is to alleviate the implications of the G-S Theorem. Crucially, this theorem itself assumes complete information. Indeed, to the best of our knowledge, there are no impossibility results regarding manipulability under incomplete information; for instance, Bayesian incentive compatibility leads to positive results [3]. So, in situations where there is complete information, assuming the ability to punish deviators implies, according to our results, that truth-telling can be enforced. On the other hand, if there is incomplete information about the preferences of the agents then the G-S Theorem does not hold, and therefore manipulation does not necessarily pose a problem. In other words, in settings where our results are irrelevant, the G-S Theorem is irrelevant as well.

Open problems. From the technical perspective the main challenge is to relax the assumptions of Theorem 4.9. It follows from the theorem's proof that, if E is stable, monotonic, and superadditive, and one constructs the GF Γ as in the proof, then the implications of the theorem hold if and only if $E_\alpha^{C(E;\cdot)} \subseteq E_*^{C(E;\cdot)}$. Hence we cannot hope to relax this assumption using our construction. However, it may certainly be the case that a different construction would require significantly weaker assumptions.

References

[1] R. J. Aumann. Acceptable points in general cooperative n-person games. In A. Tucker and R. Luce, editors, *Contributions to the Theory of Games*, volume 4, pages 287–324. Princeton University Press, 1959.

[2] R. J. Aumann. Acceptable points in games of perfect information. *Pacific Journal of Mathematics*, 10(2):381–417, 1960.

[3] C. d'Aspremont and B. Peleg. Ordinal Bayesian incentive compatible representations of committees. *Social Choice and Welfare*, 5:261–279, 1988.

[4] A. Gibbard. Manipulation of voting schemes. *Econometrica*, 41:587–602, 1973.

[5] D. Monderer and M. Tennenholtz. Strong mediated equilibrium. *Artificial Intelligence*, 173(1):180–195, 2009.

[6] M. J. Osborne and A. Rubinstein. *A Course in Game Theory.* MIT Press, 1994.

[7] P. K. Pattanaik. Counter-threats and strategic manipulation under voting schemes. *The Review of Economics Studies*, 43(1):11–18, 1976.

[8] P. K. Pattanaik. Threats, counter-threats, and strategic voting. *Econometrica*, 44(1):91–103, 1976.

[9] B. Peleg. *Game Theoretical Analysis of Voting in Committees.* Cambridge University Press, 1984.

[10] B. Peleg. Effectivity functions, game forms, games, and rights. *Social Choice and Welfare*, 15:67–80, 1998.

[11] B. Peleg and A. D. Procaccia. Implementation by mediated equilibrium. *International Journal of Game Theory*, 39(1–2):191–207, 2010.

[12] O. Rozenfeld and M. Tennenholtz. Group dominant strategies. In *Proceedings of the 3rd International Workshop on Internet and Network Economics (WINE)*, pages 457–468, 2007.

[13] O. Rozenfeld and M. Tennenholtz. Routing mediators. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1488–1493, 2007.

[14] M. Satterthwaite. Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory*, 10:187–217, 1975.