Int J Game Theory (2010) 39:191–207
DOI 10.1007/s00182-009-0175-4

Implementation by mediated equilibrium

Bezalel Peleg · Ariel D. Procaccia

Accepted: 20 August 2009 / Published online: 2 September 2009

Abstract Implementation theory tackles the following problem given a social choice correspondence (SCC), find a decentralized mechanism such that for every constellation of the individuals' preferences, the set of outcomes in equilibrium is exactly the set of socially optimal alternatives (as specified by the correspondence). In this paper we are concerned with implementation by mediated equilibrium; under such an equilibrium, the players' strategies can be coordinated in a way that discourages deviation. Our main result is a complete characterization of SCCs that are implementable by mediated strong equilibrium. This characterization, in addition to being strikingly concise, implies that some important SCCs that are not implementable by strong equilibrium are in fact implementable by mediated strong equilibrium.

Keywords Social choice · Implementation · Strong equilibrium · Mediators · Effectivity functions · Game forms

1 Introduction

A social choice correspondence (SCC) is a mapping from the preferences of individuals in a society to subsets of optimal social alternatives. A SCC gives a centralized representation of the society's morals, but in practice directly eliciting the individuals' preferences may lead to lying and manipulation.

B. Peleg
Institute of Mathematics and Center for the Study of Rationality, The Hebrew University of Jerusalem, 91904 Jerusalem, Israel
e-mail: pelegba@math.huji.ac.il

A. D. Procaccia (✉)
School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA
e-mail: arielpro@seas.harvard.edu

1.1 Implementation theory

Having in mind a specific SCC, the social planner might wish for a decentralized mechanism (formally, a *game form*) that gives rise to the same set of outcomes as the SCC, while allowing for the individuals' strategic behavior. The implementation problem can be described as follows: given a SCC, find a game form such that for any preference profile, the game's outcomes in equilibrium are exactly the socially optimal alternatives. Such a game form, which specifies the individuals' strategy spaces and the outcome given every combination of strategies, is said to *implement* the given SCC.

As is common in game theory, different equilibrium concepts can be used to capture the nature of the individuals' strategic reasoning. The implementation problem was first introduced by Maskin (1999) (although early papers by Hurwicz (1960, 1972) laid the foundations), who considered the obvious candidate: Nash equilibrium. Maskin demonstrated that two properties (of SCCs) are sufficient for implementation by Nash equilibrium: monotonicity and No Veto Power. A second prominent achievement, in the context of implementation by Nash equilibrium, is the necessary and sufficient condition (strong monotonicity) put forward by Danilov (1992).

Some research has also been devoted to implementation by strong equilibrium. Under strong equilibrium, no coalition of players is motivated to deviate in a way that benefits all its members. This line of research was again initiated by Maskin (1979), who proved that monotonicity is a necessary condition for implementability by strong equilibrium. Moulin and Peleg (1982) introduced the concept of effectivity functions, which describe the distribution of power among the individuals in a society, and used this notion to provide sufficient conditions for implementability. Dutta and Sen (1991), and later Fristrup and Keiding (2001), gave complete characterizations.

1.2 Mediated equilibria

Mediated Equilibria were first introduced by Monderer and Tennenholtz (2009), as a solution concept for games in normal form; this concept is strongly related to Aumann's c-acceptable points (Aumann 1959). Under mediated strong equilibria, the players may choose to give a mediator the right of play. The mediator then proceeds to set the empowering players' strategies; the exact choice of strategies depends on the identity of the players who have chosen to use the mediator's services. The idea is that, in case a coalition decides not to give the mediator to right of play, the mediator can set the other players' strategies in a way that punishes the rebellious coalition.

Rozenfeld and Tennenholtz (2007) considered, again in the context of games in normal form, mediators with different levels of available information. In particular, it is possible to imagine mediators that are fully aware of the strategies of the players who have *not* chosen to give them the right of play. This situation might arise, for example, in routing domains.

Peleg and Procaccia (2007) applied the ideas behind mediated equilibria to game forms. In the spirit of Rozenfeld and Tennenholtz (2007), we distinguished between two types of mediated strong equilibria: *simple* mediated strong equilibria, where each coalition has a strategy such that no matter how the other players play, they cannot improve the outcome; and *informed* mediated strong equilibria, where every coalition can respond to the strategies of the other players in a way that guarantees that the other players do not obtain a better outcome. Note that informed mediated strong equilibria are also closely related to equilibria of type II of Pattanaik (1976a). We proceeded to design social choice functions with the property that truthtelling is always a strong mediated equilibrium.

1.3 Our approach and results

We explore the power of implementation by the two types of mediated equilibria: simple/informed mediated strong Equilibria (SMSE and IMSE). Our results suggest that mediators can be quite powerful. We present two characterization of implementable SCCs, the first of which being strikingly simple compared to characterizations of SCCs that are implementable by strong equilibrium. Furthermore, our characterizations imply that important SCCs, such as the Pareto correspondence, are implementable by IMSE and not by strong equilibrium.

1.4 Structure of the paper

In Sect. 2 we give some preliminary definitions and notations. In Sect. 3 we reintroduce mediated strong equilibrium; we further discuss this concept and its relation to implementation theory. In Sect. 4, we present our main results. In Sect. 5, we study the relation between implementation by strong equilibria and implementation by mediated strong equilibria. We conclude in Sect. 6.

2 Preliminaries

In this section we elaborate on some notations and definitions that will be required in this paper. A more detailed discussion of these notions can be found in the book of Peleg (1984).

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. Throughout this paper, we deal with a finite set of players $N = \{1, 2, \ldots, n\}$, and a finite (unless explicitly stated otherwise) set of alternatives $A = \{x_1, \ldots, x_m\}$. Each player $i \in N$ holds a quasi-order R^i over A, i.e., R^i is a binary relation over A that satisfies reflexivity, antisymmetry, transitivity and totality. We let P^i be the strict preference relation associated with R^i: xP^iy iff xR^iy and $x \neq y$. The set $L = L(A)$ is the set of all such (linear) quasi-orders, so for all $i \in N$, $R^i \in L$ throughout. A *preference profile* R^N is a vector $\langle R^1, \ldots, R^n \rangle \in L^N$. We sometimes use R^S to denote the preferences of a coalition $S \in \mathcal{P}_0(N)$; xR^Sy means that xR^iy for all $i \in S$. Similarly, xP^Sy means that xP^iy for all $i \in S$. In addition, given $a \in A$, we denote the lower contour set at a according to player i by $L(a, R^i) = \{x \in A : aR^ix\}$.

A *social choice correspondence*, in its basic form, is a function $H : L^N \to \mathcal{P}_0(A)$, which maps the preferences of the voters to a desirable nonempty set of alternatives. A *social choice function* (SCF) is a function $F : L^N \to A$. In some cases we shall discuss SCCs whose domain is restricted to a set $\mathcal{D} \subseteq L^N$, i.e., functions $H : \mathcal{D} \to \mathcal{P}_0(A)$.

In our investigation we shall require some properties of SCCs. Informally, *attainability* is the set-valued equivalent of surjectivity: for every alternative there is some profile such that the alternative is in the image of the profile. H is *Maskin monotonic* if improving the position of a winning alternative does not hurt it. Finally, H is Pareto optimal if an alternative that is less preferred than another (fixed) alternative by all the agents cannot be a winner. Formally:

Definition 2.1 Let $H : \mathcal{D} \to \mathcal{P}_0(A)$, $\mathcal{D} \subseteq L^N$.

1. H is *attainable* iff for every $a \in A$ there exists $R^N \in \mathcal{D}$ such that $H(R^N) = a$.
2. H is *Maskin monotonic* iff for all $R^N, Q^N \in \mathcal{D}$, $a \in H(R^N)$,

$$\left[\forall i \in N,\ L\left(a, R^i\right) \subseteq L(a, Q^i)\right] \Rightarrow a \in H\left(Q^N\right).$$

3. H is *Pareto optimal* iff for all $x, y \in A$, $R^N \in \mathcal{D}$,

$$\left[\forall i \in N,\ xP^i y\right] \Rightarrow y \notin H\left(R^N\right).$$

2.1 Game forms

In order to formalize important game theoretic ideas, we require the notion of game forms. This notion will be readily understandable to readers familiar with very basic game theory, as a game form is simply a normal-form game stripped of the agents' payoffs. Instead, the result of a given strategy profile is one of the alternatives in A.

Definition 2.2 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 player i, and π is the *outcome function*.

Example 2.3 (King Maker game) Let $\Sigma^1 = \{2, 3\}$, and $\Sigma^2 = \Sigma^3 = A = \{a, b, c\}$. The outcome function π is given by:

$$\pi(i, x, y) = \begin{cases} x & i = 2 \\ y & i = 3 \end{cases}$$

Less formally, player 1 is the "king maker", deciding between players 2 and 3. The designated king then chooses the outcome among the three alternatives in A.

In order to obtain a true game, one has to bring into the equation incentives as well. That is, we shall consider a game to be a GF coupled with a preference profile.

Definition 2.4 Let $\Gamma = \left\langle \Sigma^1, \ldots, \Sigma^n; \pi \right\rangle$ be a GF, and let $R^N \in L^N$. The *game associated with* Γ *and* R^N is the n-person game in normal form

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

Now we can redefine some well-known solution concepts in a way that is consistent with our (abstract) notion of a game; these concepts should also be familiar to readers with a basic knowledge of game theory.

The concept of *Nash equilibrium* (1950) is perhaps the most important concept in game theory. A Nash equilibrium is a strategy profile such that no player can gain by unilaterally deviating, given that the strategies of the other players stay fixed. This provides a basic, reasonable idea of stability when the agents are rational and cannot communicate with each other.

Aumann (1959) suggested a more powerful notion of stability. A *strong equilibrium* is a strategy profile such that no *coalition* of players has an incentive to unilaterally deviate. More precisely, for every coalition of agents and every possible deviation by the coalition, there is a member of the coalition that does not gain from its participation in the deviation. Naturally, due to their restrictiveness, strong equilibria exist in preciously few settings. We now formalize the above discussion.

Definition 2.5 Let $\Gamma = \langle \Sigma^1, \ldots, \Sigma^n; \pi \rangle$ be a GF, and let $R^N \in L^N$.

1. $\sigma^N \in \Sigma^N$ is a *Nash equilibrium (NE) point* of $g(\Gamma, R^N)$ if for every $i \in N$ and every $\tau^i \in \Sigma^i$, $\pi(\sigma^N) R^i \pi(\tau^i, \sigma^{N \setminus \{i\}})$.
2. $\sigma^N \in \Sigma^N$ is a *strong equilibrium (SE) point* of $g(\Gamma, R^N)$ if for every $S \in \mathcal{P}_0(N)$ and every $\tau^S \in \Sigma^S$ there exists a player $i \in S$ such that $\pi(\sigma^N) R^i \pi\left(\tau^S, \sigma^{N \setminus S}\right)$.

2.2 Implementation

As discussed above, an SCC is a mapping from the preferences of individuals in a society to subsets of optimal social alternatives. The problem of implementation is defined thus: given an SCC, design a game form such that for any preference profile, the game's outcomes in equilibrium are exactly the alternatives selected by the SCC.

Denote the set of Nash equilibrium points of the game (Γ, R^N) by $\mathrm{NE}(\Gamma, R^N)$, and the set of strong equilibrium points by $\mathrm{SE}(\Gamma, R^N)$. Furthermore, for a set $K \subseteq \Sigma^N$, denote $\pi(K) = \{a \in A : \exists \sigma^N \in K \ s.t.\ \pi(\sigma^N) = a\}$.

Definition 2.6 The GF $\Gamma = \langle \Sigma^1, \ldots, \Sigma^b; \pi \rangle$ *implements* the SCC $H : \mathscr{D} \to R^N$, $\mathscr{D} \subseteq L^N$, by NE (resp. SE) iff for all $R^N \in \mathscr{D}$, $\pi(\mathrm{NE}(\Gamma, R^N)) = H(R^N)$ (resp. $\pi(\mathrm{SE}(\Gamma, R^N)) = H(R^N)$). H is *implementable* by NE (resp. by SE) if there exists a GF that implements H by NE (resp. SE).

Example 2.7 Let Γ be the King Maker game given in Example 2.3. Consider the SCC defined by $H(R^N) = \{t_1(R^2), t_1(R^3)\}$ for all $R^N \in L^N$, where $t_j(R)$ is the alternative ranked in place j according to R. We claim that Γ implements H by NE.

Indeed, let $R^N \in L^N$. Let $\sigma^N = \langle i, x, y\rangle$ be a NE of (Γ, R^N). If $\pi(\sigma^N) \neq t_1(R^i)$, i would want to deviate. This shows that $\pi(\text{NE}(\Gamma, R^N)) \subseteq H(R^N)$. Conversely, without loss of generality, the strategy profile $\sigma^N = \langle 2, t_1(R^2), t_3(R^1)\rangle$ is a NE of (Γ, R^N) with outcome $t_1(R^2)$. Consequently, $H(R^N) \subseteq \pi(\text{NE}(\Gamma, R^N))$.

2.3 Effectivity functions

Effectivity functions are a mathematical construct introduced by Moulin and Peleg (1982). An effectivity function, abstractly, represents the power distribution among individuals in a society. Such functions map coalitions of players to sets of subsets of alternatives. If a subset of alternatives $B \in \mathcal{P}_0(A)$ satisfies $B \in E(S)$, where E is an effectivity function, we say that S is *effective* for B. Conceptually, this means that the agents in B can force the outcome to be one of the alternatives in B. Effectivity functions have proven useful in the context of Implementation Theory.

Definition 2.8 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 $B \in \mathcal{P}_0(A)$, $B \in E(N)$.

Different notions of what it means to "force the outcome" induce different EFs. In this paper, we will deal with only three EFs; we first define two of them. α-effectiveness implies that the players in S can coordinate their strategies such that, no matter what the other players do, the outcome will be in B. If S is β-effective for B, the players in S can counter any action profile of $N\backslash S$ with actions of their own such that the outcome is in B. Clearly α-effectivity is stronger than β-effectivity.

Definition 2.9 Let $\Gamma = \langle \Sigma^1, \ldots, \Sigma^n; \pi\rangle$ be a GF, $S \in \mathcal{P}_0(N)$, $B \in \mathcal{P}_0(A)$.

1. *S is α-effective for B* if there exists $\sigma^S \in \Sigma^S$ such that for all $\tau^{N\backslash S} \in \Sigma^{N\backslash S}$, $\pi\left(\sigma^S, \tau^{N\backslash S}\right) \in B$.
2. *S is β-effective for B* if for every $\tau^{N\backslash S} \in \Sigma^{N\backslash S}$ there exists $\sigma^S \in \Sigma^S$ such that $\pi\left(\sigma^S, \tau^{N\backslash S}\right) \in B$.

We now define the α-effectivity and β-effectivity functions associated with a game form in the obvious way.

Definition 2.10 Let $\Gamma = \langle \Sigma^1, \ldots, \Sigma^n; \pi\rangle$ be a GF such that π is onto A. The *α-EF associated with* Γ is given by

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

The *β-EF associated with* Γ is given by

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

Example 2.11 Let Γ be the King Maker game given in Example 2.3, and denote $E = E_\alpha^\Gamma$. For any player $i \in N$, it holds that $E(\{i\}) = \{A\}$. On the other hand, for all $S \in \mathcal{P}_0(N)$ such that $|S| \geq 2$, $E(S) = \mathcal{P}_0(A)$, i.e. S is effective for any subset

$B \in \mathcal{P}_0(A)$. Indeed, say the coalition $\{1, 2\}$ wants to force the outcome to be a; then player 1 would choose player 2, and player 2 would choose a. Alternative a would be chosen regardless of player 3's action. We invite the reader to compute E_β^Γ.

Ironically, the third effectivity function we shall consider here is called the *first* effectivity function. Given an SCC H, a coalition S is *winning* for a subset of alternatives B if S can force the set of outcomes to be contained in B by placing B at the top of their votes. The *first effectivity function* associates with a coalition S all the subsets of alternatives for which S is winning.

Definition 2.12 Let $H : L^N \to \mathcal{P}_0(A)$ be an attainable SCC, $S \in \mathcal{P}_0(N)$, $B \in \mathcal{P}_0(A)$. S is *winning* for B iff for all $R^N \in L^N$,

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

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

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

The next definition introduces some useful properties of effectivity functions which we shall require later. An EF E is *monotonic with respect to the agents* if adding more agents to a coalition can only increase its power, and *monotonic with respect to the alternatives* if, given that a coalition is effective for a set of alternatives, it is also effective for any superset. E is *superadditive* if, given that one coalitions is effective for a subset and a second coalition is effective for another subset, then the union of the two coalitions can force an outcome in the intersection of the two subsets. E is *maximal* if, given that a coalition is not effective for a subset, then its complement is effective for the complement of the subset.

Definition 2.13 Let $E : \mathcal{P}_0(N) \to \mathcal{P}(\mathcal{P}_0(A))$.

1. E is *monotonic with respect to the players* iff for every $S \in \mathcal{P}_0(N)$ and $B \in E(S)$, if $S \subset T$ then $B \in E(T)$.
2. E is *monotonic with respect to the alternatives* iff for every $S \in \mathcal{P}_0(N)$ and $B \in E(S)$, if $B \subset B^*$ then $B^* \in E(S)$.
3. E is *monotonic* iff it is monotonic with respect to both players and alternatives.
4. E is *superadditive* iff 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 $B_1 \cap B_2 \in E(S_1 \cup S_2)$.
5. E is *maximal* iff for every $S \in \mathcal{P}_0(N)$ and $B \in \mathcal{P}_0(A)$, if $B \notin E(S)$ then $A \backslash B \in E(N \backslash S)$.

The following definition, of the core of an effectivity function, is perhaps the central definition of this section, and borrows from the same intuitions which motivate the core of a cooperative game. Say that there exists a coalition S such that all its members prefer any alternative in the subset B to x, and in addition S can force the outcome to be in B; then x cannot be a stable outcome, as the agents in S would surely gain by deviating. The *core* of E is the set of alternatives that are not unstable in this sense.

Definition 2.14 Let $E : \mathcal{P}_0(N) \to \mathcal{P}(\mathcal{P}_0(A))$, $R^N \in L^N$, $x \in A$, $S \in \mathcal{P}_0(N)$, and $B \in \mathcal{P}_0(A \backslash \{x\})$. *B dominates x via S* 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 $C(E; R^N)$.

Example 2.15 Once again, let Γ be the King Maker game given in Example 2.3, and consider the preference profile:

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

Then $C\left(E_\alpha^\Gamma; R^N\right) = \{a\}$, as $\{a\}$ dominates b and c via the coalition $S = \{1, 2\}$: the players in S both prefer a to b or c, and S is effective for $\{a\}$ (see Example 2.11).

Finally, an effectivity function is *stable* if its core is nonempty for any given profile.

Definition 2.16 An EF $E : \mathcal{P}_0(N) \to \mathcal{P}(\mathcal{P}_0(A))$ is *stable* if for all $R^N \in L^N$, $C(E; R^N) \neq \emptyset$.

3 Mediated equilibria

In previous work, Peleg and Procaccia (2007) have defined *mediated equilibria* in GFs. The idea behind mediated equilibria is that the players can commit to a course of action. In particular, the players may commit to punishing deviating coalitions, thus discouraging deviations from the predetermined outcome.

These ideas first appeared in Aumman's work on *acceptable* strategy profiles (Aumann 1959). Informally, a strategy profile is acceptable if a deviating coalition can always be punished by its complement. However, Aumann was thinking of infinite-horizon *repeated games*, where the punishment can be exacted in a future time. This reasoning is unsuitable in our setting.

Mediation is just one possible interpretation of the ability to punish deviators. The mediator is configured by the players or by another interested party, and plays only for the players who give it the right of play. Other players, who do not choose to use the mediator's services, know how the mediator is going to play for the players who do. This potentially aligns the incentives of all players with the option to empower the mediator on their behalf, and leads to a mediated equilibrium.

The term *mediated equilibrium* was first used by Monderer and Tennenholtz (2009) in the context of games in normal form. The distinction between mediated equilibria and acceptable strategies is subtle: in the Monderer–Tennenholtz setting choosing the mediator is an explicit action that can be in the support of a mixed strategy. This distinction does not survive the transition to game forms, where only pure strategies are considered.

In the spirit of Rozenfeld and Tennenholtz (2007), we distinguish between two levels of information available to the punishing coalition. Under *Simple* mediated equilibrium, the punishment is based only on the identities of the deviators. Under *informed* mediated equilibrium, the punishment can also be based on the actions of the deviating coalition. The latter situation arises, for example, in a communication network: the players must pass messages; their strategies are the routes. A router is informed of the strategies of the players, and therefore can act as a mediator (Rozenfeld and Tennenholtz 2007).

The reader might be concerned that the interpretation of the ability to punish via mediators may be inconsistent with the basic idea behind implementation theory, that is, reaching the socially desirable alternatives in a decentralized way. We argue that this is not the case. For instance, consider once again the routing domain mentioned above. The router can be configured locally in a fully decentralized way, without the intervention of the social planner, i.e., the designer of the entire network. In addition, as we have noted above, there are other interpretations of the ability to punish deviators.

Indeed, a second interpretation of punishment is the idea of *threats and counter-threats*, due to Pattanaik (1976a,b). The author formulates his concepts in the context of voting; he argues that in some situations, when a coalition of deviators forms, this is known to the complement coalition (for example, when the voting is public), and it reacts accordingly. Thus, the complement responds to the deviators' threat with a counter-threat. One of Pattainaik's solution concepts is essentially a special case (in the voting setting) of our informed mediated equilibrium.

We presently turn to our formal definitions.

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

1. $\sigma^N \in \Sigma^N$ is a *simple mediated strong equilibrium (SMSE)* point of $g\left(\Gamma, R^N\right)$ iff

$$\forall S \in \mathcal{P}_0(N) \exists \tau^S \in \Sigma^S \; s.t. \; \forall \tau^{N \setminus S} \in \Sigma^{N \setminus S} \exists i \in N \setminus S \; s.t. \; \pi\left(\sigma^N\right) R^i \pi\left(\tau^N\right).$$

2. $\sigma^N \in \Sigma^N$ is an *informed mediated strong equilibrium (IMSE)* point of $g\left(\Gamma, R^N\right)$ iff

$$\forall S \in \mathcal{P}_0(N), \forall \tau^{N \setminus S} \in \Sigma^{N \setminus S}, \; \exists \tau^S \in \Sigma^S, i \in N \setminus S \; s.t. \; \pi\left(\sigma^N\right) R^i \pi\left(\tau^N\right).$$

In the above definition, $N \setminus S$ is the deviating coalition, and S is the punishing coalition. One limitation of this definition is that we do not require the threat of the punishing coalition to be *credible*, that is, given that $N \setminus S$ indeed deviated, some of the players in the punishing coalition S may lose by carrying out their punishment. However, the underlying assumption is that players can *commit* to a course of action (e.g., via a mediator), hence the deviators know for a certainty that they will be punished, and this knowledge prevents them from deviating in the first place.

The following basic characterization result is true for strong mediated equilibria.

Lemma 3.2 (Peleg and Procaccia 2007, Lemma 3.3) *Let $\Gamma = \langle \Sigma^1, \ldots, \Sigma^n; \pi \rangle$ be a GF such that π is onto A. Then for all $R^N \in L^N$,*

1. $\pi\left(SMSE(\Gamma, R^N)\right) = C\left(E^\Gamma_\beta, R^N\right)$.
2. $\pi\left(IMSE(\Gamma, R^N)\right) = C\left(E^\Gamma_\alpha, R^N\right)$.

In the following we give a simple and direct example of implementation by mediated equilibrium.

Example 3.3 (Implementation by IMSE) Let P be the Pareto correspondence given by

$$P\left(R^N\right) = \left\{x \in A : \nexists y \in A\ s.t.\ \forall i \in N,\ yP^i x\right\}.$$

Let Γ be the "Modulo Game", defined as follows. For all $i \in N$, $\Sigma^i = A \times \{1, \ldots, n\}$, i.e., each player i picks (x^i, t^i), where $x^i \in A$ and $t^i \in \{1, \ldots, n\}$. The outcome is defined to be x^j, where $j \in N$ is the unique player satisfying $j \equiv \sum_{i \in N} t^i \pmod n$.

We claim that Γ implements P by IMSE. Indeed, let $R^N \in L^N$. If $x \notin P(R^N)$, then the grand coalition benefits by deviating; this shows that x cannot be the outcome of an IMSE.

Conversely, assume $x \in P(R^N)$. For all $S \in \mathcal{P}_0(A)$ such that $S \neq N$, and for all $\sigma^S \in \Sigma^S$, there exists $\tau^S \in \Sigma^{N\backslash S}$ such that $\pi(\sigma^S, \tau^{N\backslash S}) = x$. This is true since $N\backslash S$ can align their integers t^i in a way that $\sum_{i \in N} t^i \pmod n \in N\backslash S$. Moreover, if $S = N$, there is a player who does not want to deviate. It follows that x is the outcome of an IMSE.

As a final remark, we note that it is natural to consider the NE versions of SMSE and IMSE, that is, simple or informed mediated equilibria in which only a single deviating player must be punished. These notions are, of course, strictly weaker than NE. However, in the context of implementation, the consideration of weaker solution concepts often does not lead to more inclusive results. The reason for this is that one requires the set of equilibria of the implementing GF to be exactly equal to the image of the given SCC (instead of, say, asking that the latter be contained in the former). Indeed, it is possible to prove that any SCC that is implementable by mediated NE is also implementable by NE, using the techniques of Danilov (1992).

4 Characterization of implementable SCCs

We now turn our attention to this paper's main result: a characterization of SCCs implementable by either simple or informed mediated strong equilibria. Section 4.1 gives a concise, but possibly hard to verify, characterization. Section 4.2 makes this characterization more tractable by further breaking down the conditions.

4.1 First (concise) characterization

We begin with implementation by SMSE. Notice that the theorems in this subsection hold for SCCs $H : \mathscr{D} \to \mathcal{P}_0(A)$, where $\mathscr{D} \subseteq L^N$ is an arbitrary domain of preference profiles.

Theorem 4.1 *Let $H : \mathscr{D} \to \mathcal{P}_0(A)$, $\mathscr{D} \subseteq L^N$, be an attainable SCC. H is implementable by simple mediated strong equilibrium if, and only if, there exists a monotonic and maximal EF $E : \mathcal{P}(N) \to \mathcal{P}(\mathcal{P}_0(A))$ such that $\forall R^N \in \mathscr{D},\ H(R^N) = C(E, R^N)$.*

Moreover, the implementing GF can be chosen to be Γ^F*, where* F *is a social choice function* $F : L^N \to A$.

The GF Γ^F, mentioned in the theorem's statement, is given by $\Gamma^F = \langle L, \ldots, L; F \rangle$; indeed, in this GF the players' strategies are orderings of alternatives, and the outcome is determined by F. Essentially, Γ^F is completely equivalent to the SCF F.

In order to prove this theorem, we require two previously known results.

Lemma 4.2 (Peleg 1984, Remarks 6.1.9 and 6.1.15) *Let* Γ *be a GF. Then* E_α^Γ *and* E_β^Γ *are monotonic, and* E_α^Γ *is superadditive.*

Lemma 4.3 (Peleg and Procaccia 2007, Theorem 4.2) *Let* $E : \mathcal{P}(N) \to \mathcal{P}(\mathcal{P}_0(A))$ *be a stable and maximal EF, and let* $F : L^N \to A$ *such that* $F(R^N) \in C(E, R^N)$ *for all* $R^N \in L^N$*. Then* $E_\beta^{\Gamma^F} = E$.

Proof of Theorem 4.1 Assume first that H is implementable by simple mediated strong equilibrium. Let $\Gamma = \langle \Sigma^1, \ldots, \Sigma^n; \pi \rangle$ be the implementing GF; we claim that E_β^Γ is as required.

We first verify that E_β^Γ is indeed an EF. Clearly, for all $S \in \mathcal{P}_0(N)$, $A \in E_\beta^\Gamma(S)$. Furthermore, since H is attainable and Γ implements H, π must be onto A. That is, for every $a \in A$ there exists $\sigma^N \in \Sigma^N$ such that $\pi(\sigma) = a$. It follows that N is β-effective for $\{a\}$ (by using σ^N). We conclude (since E_β^Γ is monotonic with respect to the alternatives by Lemma 4.2) that $E_\beta^\Gamma(N) = \mathcal{P}_0(A)$.

Now, we have that for all $R^N \in \mathscr{D}$,

$$H\left(R^N\right) = \text{SMSE}\left(\Gamma, R^N\right) = C\left(E_\beta^\Gamma, R^N\right), \tag{1}$$

where the first equality is true as Γ is an implementation of H by SMSE, and the second equality follows from Lemma 3.2.

Finally, we must show that E_β^Γ is maximal. Let $S \in \mathcal{P}_0(N)$, $B \in \mathcal{P}_0(A)$. If $B \notin E_\beta^\Gamma$, then there exists $\tau^{N\setminus S} \in \Sigma^{N\setminus S}$ such that for all $\sigma^S \in \Sigma^S$, $\pi(\sigma^S, \tau^{N\setminus S}) \in A\setminus B$. Thus, $A\setminus B \in E_\alpha^\Gamma(N\setminus S)$, and in particular $A\setminus B \in E_\beta^\Gamma(N\setminus S)$.

Conversely, assume that there exists a maximal and stable EF E such that $\forall R^N \in \mathscr{D},\ H(R^N) = C(E, R^N)$. Let $H^* : L^N \to \mathcal{P}_0(A)$ be the extension of H to L^N such that $\forall R^N \in L^N,\ H^*(R^N) = C(E, R^N)$. Let $F : L^N \to A$ such that $F(R^N) \in C(E, R^N)$ for all $R^N \in L^N$. By Lemma 4.3, $E_\beta^{\Gamma^F} = E$. Therefore, for all $R^N \in L^N$,

$$H^*(R^N) = C(E, R^N) = C\left(E_\beta^{\Gamma^F}, R^N\right) = \text{SMSE}\left(\Gamma^F, R^N\right),$$

where the first equality follows from the assumption, the second by the above-mentioned theorem, and the third equality is a consequence of Lemma 3.2. In particular, for all $R^N \in \mathscr{D}$,

$$H\left(R^N\right) = H^*\left(R^N\right) = \text{SMSE}\left(\Gamma^F, R^N\right). \qquad \square$$

Remark 4.4 It is possible to drop the monotonicity of E from the characterization. We leave it in as it provides a unified interface for Theorems 4.1 and 4.5, which will later enable us to plug in our next characterization.

The characterization of implementation by IMSE is quite similar, the only difference being that the maximality of E (which is not a weak requirement) is replaced by superadditivity (which is).

Theorem 4.5 *Let $H : \mathscr{D} \to \mathcal{P}_0(A)$, $\mathscr{D} \subseteq L^N$, be an attainable SCC. H is implementable by informed mediated strong equilibrium if, and only if, there exists a monotonic and superadditive EF $E : \mathcal{P}(N) \to \mathcal{P}(\mathcal{P}_0(A))$ such that $\forall R^N \in \mathscr{D}$, $H(R^N) = C(E, R^N)$.*

We require the following additional lemma.

Lemma 4.6 (Peleg 1998, Theorem 3.5) *Let $E : \mathcal{P}(N) \to \mathcal{P}(\mathcal{P}_0(A))$ be an EF. Then E is monotonic and superadditive if, and only if, there exists a GF Γ such that $E = E_\alpha^\Gamma$.*

Proof of Theorem 4.5 Assume first that H is implementable by informed mediated strong equilibrium. Let Γ be the implementing GF; we will show that E_α^Γ is as required. As in the proof of Theorem 4.1, E_α^Γ is an EF due to the attainability assumption. For all $R^N \in \mathscr{D}$ it holds that

$$H(R^N) = \text{IMSE}\ (\Gamma, R^N) = C\left(E_\alpha^\Gamma, R^N\right). \tag{2}$$

The first equality follows from the fact that Γ is an implementation of H by IMSE; the second equality is implied by Lemma 3.2. By Lemma 4.2, E_α^Γ is monotonic and superadditive.

In the other direction, let E be a monotonic and superadditive EF such that $\forall R^N \in \mathscr{D}$, $H(R^N) = C(E, R^N)$. By Lemma 4.6, since E is monotonic and superadditive, there exists a GF Γ such that $E = E_\alpha^\Gamma$; we claim that the foregoing GF Γ implements H by IMSE. Indeed, we have that for all $R^N \in \mathscr{D}$,[1]

$$H\left(R^N\right) = C\left(E, R^N\right) = C\left(E_\alpha^\Gamma, R^N\right) = \text{IMSE}\left(\Gamma, R^N\right),$$

where the first equality follows from the assumption, the second is a consequence of the construction of Γ, and the third is implied by Lemma 3.2. □

4.2 Second (tractable) characterization

Although Theorems 4.1 and 4.5 give concise necessary and sufficient conditions for implementation by SMSE and IMSE, respectively, these conditions may still be hard to verify. The main problem is that the conditions ask for the *existence* of an EF with

[1] As in the proof of Theorem 4.1, one can explicitly define H^* as the extension of H to L^N, but this is not a mathematical necessity but rather a pedagogical tool.

certain properties. We would like to obtain a more tractable characterization, via the observation that this EF must be chosen to be $E^*(H)$. Indeed, the following lemma is previously known.

Lemma 4.7 (Peleg 1984, Lemma 6.1.21) *Let $E : \mathcal{P}(N) \to \mathcal{P}(\mathcal{P}_0(A))$ be a stable and monotonic function. If $H(R^N) = C(E, R^N)$ for every $R^N \in L^N$, Then $E^*(H) = E$.*

Our next theorem yields more tractable, albeit less concise, characterizations as an easy corollary. In contrast to Subsection 4.1, heretofore the results are formulated for SCCs whose domain is the universal domain L^N. We shall need the following definition (Peleg 1984, Definition 3.2.4): An SCC $H : L^N \to \mathcal{P}_0(A)$ is *core-inclusive* with respect to a function $E : \mathcal{P}_0(A) \to \mathcal{P}(\mathcal{P}_0(A))$ iff for all $R^N \in L^N$, $C(E, R^N) \subseteq H(R^N)$.

Theorem 4.8 *Let $H : L^N \to \mathcal{P}_0(A)$. There exists a monotonic EF $E : \mathcal{P}(N) \to \mathcal{P}(\mathcal{P}_0(A))$ such that $\forall R^N \in L^N,\ H(R^N) = C(E, R^N)$ if, and only if, the following conditions hold:*

1. *H is Pareto optimal.*
2. *H is Maskin Monotonic.*
3. *H is core-inclusive with respect to $E^* = E^*(H)$.*

In order to prove the theorem, we require several additional lemmata.

Lemma 4.9 (Peleg 1984, Remark 5.3.12) *Let $E : \mathcal{P}(N) \to \mathcal{P}(\mathcal{P}_0(A))$ be a function. Then $C(E, \cdot)$ is Maskin monotonic.*

Lemma 4.10 (Peleg 1984, Lemma 6.1.20) *Let $E : \mathcal{P}(N) \to \mathcal{P}(\mathcal{P}_0(A))$ be a stable function, and let $H(R^N) = C(E, R^N)$ for every $R^N \in L^N$. Then $E^*(H)$ is monotonic.*

Lemma 4.11 (Peleg 1984, Lemma 6.5.6) *Let $H : L^N \to \mathcal{P}_0(A)$. If H is Maskin monotonic then for all $R^N \in L^N$, $H(R^N) \subseteq C(E^*(H), R^N)$.*

Proof of Theorem 4.8 Assume that there exists a monotonic EF $E : \mathcal{P}(N) \to \mathcal{P}(\mathcal{P}_0(A))$ such that $\forall R^N \in L^N,\ H(R^N) = C(E, R^N)$. We first prove condition 1, namely Pareto optimality. Let $x, y \in A$ and $R^N \in L^N$ such that for all $i \in N$, $x P^i y$. Since E is an EF, $\{x\} \in E(N)$. Therefore, $\{x\}$ dominates y via N, i.e., $y \notin C(E, R^N) = H(R^N)$.

Now, condition 2 is readily satisfied by Lemma 4.9. Moreover, By Lemma 4.7, $E = E^*$. Therefore, $H(R^N) = C(E^*, R^N)$ for all $R^N \in L^N$, and in particular H is core-inclusive with respect to E^* (i.e., condition 3 is satisfied as well).

Conversely, assume conditions 1–3 hold. We will show that E^* is as required. By Lemma 4.11, $H(R^N) \subseteq C(E^*, R^N)$ for all $R^N \in L^N$, and together with the assumption that H is core-inclusive with respect to E^* we obtain that $\forall R^N \in L^N,\ H(R^N) = C(E^*, R^N)$. Now, by Lemma 4.10, E^* is monotonic.

We argue that since H is Pareto optimal, E^* is an EF. Clearly for all $S \in \mathcal{P}_0(N)$, $A \in E(S)$. Moreover, let $a \in A$; let $R^N \in L^N$ such that all players rank a first. By Pareto optimality, $H(R^N) = \{a\}$. In other words, $\{a\} \in E^*(N)$. We now have that $E^*(N) = \mathcal{P}_0(A)$ as E^* is monotonic with respect to the alternatives. □

We can now give a second characterization of SCCs that are implementable by mediated equilibria, by combining Theorems 4.1, 4.5, and 4.8.

Corollary 4.12 *Let $H : L^N \to \mathcal{P}_0(A)$, be an attainable SCC*

1. *H is implementable by simple mediated strong equilibrium if, and only if, Theorem 4.8's conditions 1–3 hold and $E^*(H)$ is maximal.*
2. *H is implementable by informed mediated strong equilibrium if, and only if, Theorem 4.8's conditions 1–3 hold and $E^*(H)$ is superadditive.*

Remark 4.13 Theorem 4.1 can be slightly strengthened, by removing the assumptions that H is attainable and that E is an EF. We say that a function $E : \mathcal{P}(N) \to \mathcal{P}(\mathcal{P}_0(A))$ is a *pseudo-EF* if for every $S \in \mathcal{P}_0(N)$, $A \in E(S)$. The following statement is true: Let $H : \mathscr{D} \to \mathcal{P}_0(A)$, $\mathscr{D} \subseteq L^N$. H is implementable by simple mediated strong equilibrium iff there exists a monotonic and maximal pseudo-EF $E : \mathcal{P}(N) \to \mathcal{P}(\mathcal{P}_0(A))$ such that $\forall R^N \in \mathscr{D},\ H(R^N) = C(E, R^N)$.

Now, Theorem 4.8 can also be modified accordingly, by abandoning the condition that H is Pareto optimal. That is, there exists a monotonic pseudo-EF $E : \mathcal{P}(N) \to \mathcal{P}(\mathcal{P}_0(A))$ such that $\forall R^N \in L^N,\ H(R^N) = C(E, R^N)$ iff H is Maskin Monotonic and H is core-inclusive with respect to $E^* = E^*(H)$.

5 The power of implementation by strong mediated equilibria

In this section we will attempt to understand the power of implementation by mediated strong equilibrium, as compared to implementation by strong equilibrium. Previous work has given complete characterizations of implementation by strong equilibrium (Dutta and Sen 1991; Fristrup and Keiding 2001). Alas, these characterizations are rather hard to formulate, if not to verify. So, one immediate advantage of implementation by mediated strong equilibria is that Theorems 4.1 and 4.5 are relatively simple.

Let us discuss implementation by SMSE. On the face of it, simple mediators cannot help. Indeed, we have the following theorem:

Theorem 5.1 *Let $H : L^N \to \mathcal{P}_0(A)$. If H is implementable by SMSE, then H is implementable by SE.*

The theorem follows almost immediately from the following lemma.

Lemma 5.2 (Peleg 1984, Theorem 6.4.2) *Let $E : \mathcal{P}_0(N) \to \mathcal{P}(\mathcal{P}_0(A))$ be a stable and monotonic EF. Then the core $C(E, \cdot)$ is implementable by SE if, and only if, E is maximal.*

Proof of Theorem 5.1 By Theorem 4.1, if H is implementable by SE, then there exists a monotonic and maximal (and stable) EF E such that H is the core of E. By Lemma 5.2 H is implementable by SE. □

Theorem 5.1 may come as an unpleasant surprise. However, notice that in certain settings, even simple mediators offer a substantial advantage over implementation by SE. Recall that Theorem 4.1 states that the implementing GF may be chosen to be a SCF. This is no small thing; the implementing GF is a description of a decentralized mechanism the agents are expected to use in order to strategically reach a collective decision. It is very significant that this GF be as simple as possible. The implementing GF given in the proof of Lemma 5.2, for instance, is less intuitive. In general, it is unknown whether implementation by SE is possible when the implementing GF is a SCF. That said, we note that the implementing GF constructed in the proof of Lemma 5.2 is defined directly from the EF E, while the one constructed in the proof of Theorem 4.1 requires the computation of the core of an EF; this task may prove intractable (Mizutani et al. 1993).

We move on to implementation by IMSE. The frequency of IMSEs, compared to SEs, is not necessarily an advantage. One would expect informed mediators to help when implementing "large" SCCs, but not when implementing "small" ones. This is indeed the case.

As a general example, consider an EF E that is monotonic, superadditive, and stable, but not maximal. By Theorem 4.5, $C(E, \cdot)$ is implementable by IMSE; by Lemma 5.2, $C(E, \cdot)$ is not implementable by SE.

We now give two specific examples: the first is a very important SCC that is implementable by IMSE and not by SE. The second, interestingly, is of a SCC (admittedly, a very nonintuitive one) that is implementable by SE and not by IMSE.

Example 5.3 (SCC that is implementable by IMSE and not by SE) Consider the prominent Pareto correspondence given by $P(R^N) = \{x \in A : \nexists y \in A\, s.t.\, \forall i \in N,\ yP^i x\}$.

In Example 3.3, we have shown that P is implementable by IMSE, and that the implementing GF can be chosen to be the modulo game. However, notice that this result also easily follows from our theorems (albeit via a more complex implementing GF). Define an EF E by $E(N) = \mathcal{P}_0(A)$, $E(S) = \{A\}$ for all $N \neq S \in \mathcal{P}_0(N)$. It is easily seen that for all R^N, $P(R^N) = C(E, R^N)$, and that E is monotonic and superadditive. By Theorem 4.5, P is implementable by IMSE.

On the other hand, it is also straightforward that E is not maximal, and thus according to Lemma 5.2 P is not implementable by SE.

Example 5.4 (SCC that is implementable by SE and not by IMSE) Let $N = \{1, 2, 3, 4\}$, $A = \{a, b, c\}$. Let $\beta_*(a) = 1$ and $\beta_*(b) = \beta(c) = 2$, and let $R^N \in L^N$. Define an EF $E_* = E_*(\beta_*)$ by:

$$B \in E_*(S) \Leftrightarrow |S| \geq \beta_*(A \setminus B),$$

where $\beta_*(B) = \sum_{x \in B} \beta_*(x)$.

Now, define $H : L^N \to \mathcal{P}_0(A)$ by the following rules. If there exists $x \in A$ such that

$$|\{i \in N : x \text{ is ranked first in } R^i\}| \geq 3,$$

then $H(R^N) = \{x\}$. Otherwise, $H(R^N) = C(E_*, R^N)$.

Further, let $F : L^N \to A$ be a selection from H, i.e., $F(R^N) \in H(R^N)$ for all $R^N \in L^N$, and let $\Gamma^F = \langle L, \ldots, L; F \rangle$ be the GF that is determined by F. Define H^* by:

$$H^*\left(R^N\right) = F\left(SE\left(\Gamma^F, R^N\right)\right)$$

for all $R^N \in L^N$.

Peleg (1984, Example 6.5.7) proves that $SE(\Gamma^F, R^N) \neq \emptyset$ for all $R^N \in L^N$, so clearly H^* is implementable by SE. Peleg also proves that H^* is *not* the core correspondence of an EF. By Theorem 4.5, H^* is not implementable by IMSE.

Finally, we observe that implementation by mediated strong equilibrium, of either type, implies implementation by Nash equilibrium when $n \geq 3$. Indeed, Peleg and Winter (2002) prove:

Lemma 5.5 (Peleg and Winter 2002, Lemma 3.3) *Let $E : \mathcal{P}_0(N) \to \mathcal{P}(\mathcal{P}_0(A))$ be a stable EF. If $n \geq 3$, then the core $C(E, R^N)$ is implementable by NE.*

Together with Theorems 4.1 and 4.5 we obtain:

Theorem 5.6 *Let $H : L^N \to \mathcal{P}_0(A)$, $n \geq 3$. If H is implementable by SMSE or IMSE, then H is implementable by NE.*

6 Conclusions

We have considered implementation by mediated equilibria. Our main result is a characterization of SCCs implementable by SMSE or IMSE. Informally, our concise characterization states that an SCC is implementable by SMSE (resp. IMSE) iff it is the core correspondence of a monotonic and maximal (resp. and superadditive) EF. Using this characterization, we have shown that any SCC implementable by SMSE is implementable by SE, but have noted an important distinction: the implementing GF in the case of SMSE can be chosen to be a SCF. Crucially, we have discussed the power of informed mediators, showing that certain SCCs (such as the important Pareto correspondence) are implementable by IMSE and not by SE.

References

Aumann RJ (1959) Acceptable points in general cooperative n-person games. In: Tucker A, Luce R(eds) Contributions to the Theory of Games, vol 4, pp 287–324. Princeton University Press

Danilov VI (1992) Implementation via Nash equilibria. Econometrica 60(1):43–56

Dutta B, Sen A (1991) Implementation under strong equilibrium: a complete characterization. J Math Econ 20:49–67

Fristrup P, Keiding H (2001) Strongly implementable social choice correspondences and the supernucleus. Soc Choice Welf 18:213–226

Hurwicz L (1960) Optimality and informational efficiency in resource allocation processes. In: Arrow KJ, Karlin S, Suppes P (eds) Mathematical methods in the social sciences. Stanford University Press, Stanford, pp 27–46

Hurwicz L (1972) On informationally decentralized systems. In: Radner R, McGuire CB (eds) Decision and organization. North Holland, Amsterdam, pp 297–336

Maskin E (1979) Implementation and strong Nash equilibrium. In: Laffont JJ (ed) Aggregation and revelation of preferences. North Holland, Amsterdam, pp 433–439

Maskin E (1999) Nash equilibrium and welfare optimality. Rev Econ Stud 66:23–38. This paper was first circulated in 1977

Mizutani M, Hiraide Y, Nishino H (1993) Computational complexity to verify the unstability of effectivity function. Int J Game Theory 22(3):225–239

Monderer D, Tennenholtz M (2009) Strong mediated equilibrium. Artif Intell 173(1):180–195

Moulin H, Peleg B (1982) Cores of effectivity functions and implementation theory. J Math Econ 10:115–145

Nash JF (1950) Equilibrium points in N-person games. Proc Natl Acad Sci USA 36:48–49

Pattanaik PK (1976a) Counter-threats and strategic manipulation under voting schemes. Rev Econ Stud 43(1):11–18

Pattanaik PK (1976b) Threats, counter-threats, and strategic voting. Econometrica 44(1):91–103

Peleg B (1984) Game theoretical analysis of voting in committees. Cambridge University Press, Cambridge

Peleg B (1998) Effectivity functions, game forms, games, and rights. Soc Choice Welf 15:67–80

Peleg B, Procaccia AD (2007) Mediators enable truthful voting. Discussion paper 451, Center for the Study of Rationality. The Hebrew University of Jerusalem

Peleg B, Winter E (2002) Constitutional implementation. Rev Econ Des 7:187–204

Rozenfeld O, Tennenholtz M (2007) Routing mediators. In: Proceedings of the 20th international joint conference on artificial intelligence (IJCAI), pp 1488–1493