האוניברסיטה העברית בירושלים
THE HEBREW UNIVERSITY OF JERUSALEM

IMPLEMENTATION BY MEDIATED EQUILIBRIUM

by

BEZALEL PELEG and ARIEL D. PROCACCIA

Discussion Paper # 463 September 2007

מרכז לחקר הרציונליות

CENTER FOR THE STUDY OF RATIONALITY

Feldman Building, Givat-Ram, 91904 Jerusalem, Israel
PHONE: [972]-2-6584135 FAX: [972]-2-6513681
E-MAIL: ratio@math.huji.ac.il
URL: http://www.ratio.huji.ac.il/

Implementation by Mediated Equilibrium

Bezalel Peleg* Ariel D. Procaccia†

September 9, 2007

Abstract

Implementation theory tackles the following problem: given a social choice correspondence, 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, a mediator coordinates the players' strategies in a way that discourages deviation. Our main result is a complete characterization of social choice correspondences which are implementable by mediated strong equilibrium. This characterization, in addition to being strikingly concise, implies that some important social choice correspondences which are not implementable by strong equilibrium are in fact implementable by mediated strong equilibrium.

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.

Implementation Theory. Having in mind a specific SCC, the social planner might wish for a decentralized mechanism (formally, a *game form*) which 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.

*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 Computer Science, The Hebrew University of Jerusalem. email: arielpro@cs.huji.ac.il

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 [12] (although early papers by Hurwicz [9, 10] 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 [4].

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 which benefits all its members. This line of research was again initiated by Maskin [11], who proved that monotonicity is a necessary condition for implementability by strong equilibrium. Moulin and Peleg [15] 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 [6], and later Fristrup and Keiding [7], gave complete characterizations.

Mediated Equilibria. Mediated Equilibria were first introduced by Monderer and Tennenholz [14], as a solution concept for games in normal form; this concept is strongly related to Aumann's c-acceptable points [3]. 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 Tennenholz [20] 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 which 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 [18] applied the ideas behind mediated equilibria to game forms. In the spirit of Rozenfeld and Tennenholz [20], Peleg and Procaccia 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. Peleg and Procaccia proceeded to design social choice functions with the property that truthtelling is always a strong mediated equilibrium.

Our approach and results. Our contribution begins by extending the definition of mediated equilibria, in the obvious way, to mediated Nash equilibria.

We then explore the power of implementation by the four types of mediated equilibria: simple/informed mediated Nash/strong Equilibria (SMNE, IMNE, SMSE, and IMSE). We show that any SCC which is implementable by SMNE or IMNE is also implementable by Nash equilibrium.

In contrast, when it comes to implementation by mediated strong equilibrium, mediators turn out to be quite powerful. We present two characterization of implementable SCCs, the first of which being strikingly simple compared to characterizations of SCCs which 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.

Structure of the paper. In Section 2 we give some preliminary definitions and notations. In Section 3 we briefly reintroduce mediated strong equilibria, and formally define mediated Nash equilibria. In Section 4, we discuss implementation by mediated Nash equilibrium. In Section 5, we investigate implementation by mediated strong equilibrium. We conclude in Section 6. Finally, in the appendix we examine the consistency of game forms with respect to mediated equilibria.

2 Preliminaries

In this section we elaborate on some notations and definitions which will be required in this paper. A more detailed discussion of these notions can be found in Peleg [16].

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 which 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* (SCC), 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 social choice correspondences whose domain is restricted to a set $\mathscr{D} \subseteq L^N$, *i.e.*, functions $H : \mathscr{D} \to \mathcal{P}_0(A)$.

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

1. H is *attainable* iff for every $a \in A$ there exists $R^N \in \mathscr{D}$ such that $H(R^N) = a$.

2. H is *Maskin monotonic* iff for all $R^N, Q^N \in \mathscr{D}$, $a \in H(R^N)$,

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

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

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

2.1 Game Forms and Implementation

Informally, a game form is a normal-form game stripped of the players' 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 game form coupled with a preference profile.

Definition 2.4. Let $\Gamma = \langle \Sigma^1, \ldots, \Sigma^n; \pi \rangle$ 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

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

Now we can redefine some well-known solution concepts in a way which is consistent with our (more abstract) notion of a game.

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 $\pi(\sigma^N) R^i \pi(\tau^S, \sigma^{N \setminus S})$.

Denote the set of Nash equilibrium points of the game (Γ, R^N) by $NE(\Gamma, R^N)$, and the set of strong equilibrium points by $SE(\Gamma, R^N)$. 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 $R^N \in \mathscr{D}$, $\pi(NE(\Gamma, R^N)) = H(R^N)$ (resp. $\pi(SE(\Gamma, R^N)) = H(R^N)$). H is *implementable* by NE (resp. by SE) if there exists a GF which 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, $H(R^N) \subseteq \pi(\text{NE}(\Gamma, R^N))$.

2.2 Effectivity Functions

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 $B \in \mathcal{P}_0(A)$ 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.

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 effectivity functions. In this paper, we will deal with only three effectivity functions; 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 $N \setminus S$ 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 $\tau^{N \setminus S} \in \Sigma^{N \setminus S}$, $\pi(\sigma^S, \tau^{N \setminus S}) \in B$.

2. *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 $\pi(\sigma^S, \tau^{N \setminus S}) \in B$.

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

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

The *β-effectivity function associated with* Γ is given by

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

Example 2.11. Let Γ be the King Maker game given in Example 2.3, and denote $E = E^{\Gamma}_{\alpha}$. 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^{Γ}_{β}.

We now define the third effectivity function we shall consider here (ironically called the *first* effectivity function).

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

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

The *first effectivity function* 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.

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 w.r.t. 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 *subadditive* iff for every $S_i \in \mathcal{P}_0(N), B_i \in E(S_i)$, $i = 1, 2$, if $B_1 \cap B_2 = \emptyset$ then $B_1 \cup B_2 \in E(S_1 \cap S_2)$.

6. 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 $A \setminus B \in E(N \setminus S)$.

The following definition, of the core of an effectivity function, borrows from the same intuitions which motivate the core of a cooperative game.

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 $B \in \mathcal{P}_0(A \setminus \{x\})$. *$B$ 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 $C(E; R^N)$.

If B dominates x via S, x is (in a sense) unstable, as the players 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.

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 $C(E^\Gamma_\alpha; R^N) = \{a\}$, 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).

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

3 Mediated Equilibria

Peleg and Procaccia [18] interpreted the presence of a mediator as an option, available to the players, to commit to a particular course of action. The mediator is configured by the players or by another interested party, and plays only for the players which give it the right of play. Other players, which 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.

Peleg and Procaccia [18], in the spirit of Rozenfeld and Tennenholz [20], distinguished between two levels of information available to the mediators. *Simple* mediators only know which players choose to use their services. *Informed mediators*, in addition to basing their actions on knowledge of the empowering players, are also aware of the strategy profile of the players who do not give them the right of play. This situation arises, for example, in a routing domain where the mediator is a router [20].

The following definition of simple/informed mediated equilibrium, taken from Peleg and Procaccia [18], abstracts away the explicit presence of a mediator. A strategy profile is a simple mediated strong equilibrium point if for every deviating coalition (*i.e.*, a coalition which does not use the mediator's services), its complement (presumably using the services of the mediator) can be configured to punish the deviators. Therefore, conceptually a mediator can be configured to play the simple mediated strong equilibrium point for all players

should everybody choose to use its services, and deter potential deviators. In an informed mediated strong equilibrium point, the punishing coalition (empowering the mediator) can base its punishing strategy on knowledge of the deviators' strategies.

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

1. $\sigma^N \in \Sigma^N$ is a *simple mediated strong equilibrium (SMSE)* point of $g(\Gamma, R^N)$ iff
$$\forall S \in 2^N \exists \tau^S \in \Sigma^S \ s.t. \ \forall \tau^{N \setminus S} \in \Sigma^{N \setminus S} \exists i \in N \setminus S \ s.t. \ \pi(\sigma^N) R^i \pi(\tau^N).$$
2. $\sigma^N \in \Sigma^N$ is an *informed mediated strong equilibrium (IMSE)* point of $g(\Gamma, R^N)$ iff
$$\forall S \in 2^N, \forall \tau^{N \setminus S} \in \Sigma^{N \setminus S}, \ \exists \tau^S \in \Sigma^S, i \in N \setminus S \ s.t. \ \pi(\sigma^N) R^i \pi(\tau^N).$$

In the above definition, $N \setminus S$ is the deviating coalition, and S is the punishing coalition, *i.e.*, the players which gave the mediator the right of play.

In this paper, we also consider mediated Nash equilibria. The definition of a *simple mediated Nash equilibrium (SMNE)* point is identical to that of SMSE, except that the mediator only has to deal with deviating coalitions $N \setminus S$ which are singletons. The same goes for the definition of an *informed mediated Nash equilibrium (IMNE)*: this too can be defined by similarly tampering with the definition of IMSE.

Definition 3.2. Let $\Gamma = \langle \Sigma^1, \ldots, \Sigma^n; \pi \rangle$ be a game form, $R^N \in L^N$.

1. $\sigma^N \in \Sigma^N$ is a *simple mediated Nash* equilibrium (SMNE) point of $g(\Gamma, R^N)$ iff
$$\forall i \in N \exists \tau^{N \setminus \{i\}} \in \Sigma^{N \setminus \{i\}} \ s.t. \ \forall \tau^i \in \Sigma^i, \ \pi(\sigma^N) R^i \pi(\tau^N).$$
2. $\sigma^N \in \Sigma^N$ is an *informed mediated Nash* equilibrium (IMNE) point of $g(\Gamma, R^N)$ iff
$$\forall i \in N, \forall \tau^i \in \Sigma^i, \ \exists \tau^{N \setminus \{i\}} \in \Sigma^{N \setminus \{i\}} \ s.t. \ \pi(\sigma^N) R^i \pi(\tau^N).$$

Conceptually, then, in the context of mediated Nash equilibria, the mediator can be configured purposefully for coalitions of size n or $n-1$ which choose to give it the right of play (in a way which punishes lone deviators), and arbitrarily for smaller coalitions.

So, we have four types of mediated equilibria: SMNE, IMNE, SMSE, and IMSE. As in the case of NE, we denote, by SMNE(Γ, R^N) (resp., IMNE(Γ, R^N), SMSE(Γ, R^N), IMSE(Γ, R^N)) the set of simple mediated Nash (resp., informed mediated Nash, simple mediated strong, informed mediated strong) equilibria in (Γ, R^N). A GF Γ implements $H : \mathscr{D} \to \mathcal{P}_0(A)$ by SMNE (resp., IMNE, SMSE, IMSE) if it holds that for all $R^N \in \mathscr{D}$, $\pi(\text{SMNE}(\Gamma, R^N)) = H(R^N)$ (resp., $\pi(\text{IMNE}(\Gamma, R^N)) = H(R^N)$, $\pi(\text{SMSE}(\Gamma, R^N)) = H(R^N)$, $\pi(\text{IMSE}(\Gamma, R^N)) = H(R^N)$).

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

Lemma 3.3. *[18, Lemma 3.3] Let $\Gamma = \langle \Sigma^1, \ldots, \Sigma^n; \pi \rangle$ be a game form such that π is onto A. Then for all $R^N \in L^N$,*

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

Before proceeding to more technical results, we would like to devote the end of the introductory part of the paper to a simple and direct example of implementation by mediated equilibrium.

Example 3.4 (Implementation by IMSE)**.** Let P be the Pareto correspondence given by $P(R^N) = \{x \in A : \nexists y \in A \ s.t. \ \forall i \in N, \ yP^i x\}$. 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 $j \equiv \sum_{i \in N} t^i (\text{mod } n)$.

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 $\tau^S \in \Sigma^{N \setminus S}$ such that $\pi(\sigma^S, \tau^{N \setminus S}) = x$. This is true since $N \setminus S$ can align their integers t^i in a way that $\sum_{i \in N} t^i (\text{mod } n) \in N \setminus S$. Moreover, if $S = N$, there is a player who does not want to deviate. It follows that x is the outcome of an IMSE.

4 Implementation by Mediated Nash Equilibrium

We begin the presentation of our contribution by establishing a surprising result regarding implementation by SMNE and IMNE. Indeed, in this section we shall prove:

Theorem 4.1. *Let $H : L^N \to \mathcal{P}_0(A)$. If H is implementable by simple/informed mediated Nash equilibrium, then H is implementable by Nash equilibrium.*

Remark 4.2. Theorem 4.1 holds even if the set of alternatives is infinite.

Clearly, any Nash equilibrium is both a SMNE and an IMNE, but the opposite is not true. Moreover, there are games where mediated Nash equilibria exist while Nash equilibria do not.

Example 4.3. We give an example of a game that has an IMNE but no SMNE (and, in particular, no NE). Consider the following "matching pennies" game; the game form Γ is given by:

	L	R
U	a	b
D	b	a

The preference profile R^N is given by: aR^1b, bR^2a. It is easily seen that (Γ, R^N) has no SMNE, and on the other hand, every strategy profile is an IMNE.

Let us extend this example to obtain a game that has a SMNE but no NE. The game form Γ is given by:

	L	M	R
U	a	b	c
M	b	a	c
D	c	c	d

The preference profile is given by: $aR^1bR^1cR^1d$, $bR^2aR^2cR^2d$. The reader is invited to verify that the game (Γ, R^N) has no NE. On the other hand, every strategy profile with outcome a or b (namely $(U, L), (U, M), (M, L), (M, M)$) is a SMNE.

However, as Theorem 4.1 implies, the prevalence of mediated equilibria is not an advantage when it comes to implementation. The reason for this is that one requires the set of equilibria of the implementing game form to be *exactly equal* to the image of the given SCC (instead of, say, asking that the latter be contained in the former).

Let us now discuss the proof of Theorem 4.1. Danilov [4] defined a strong notion of monotonicity. If H is a SCC, we say that $a \in A$ is *essential* for $i \in N$ in $B \in \mathcal{P}_0(A)$ if there exists $R^N \in L^N$ such that $a \in H(R^N)$ and $L(a, R^i) \subseteq B$. Denote:

$$Ess_i(B) = \{a \in B : \ a \text{ is essential for } i \text{ in } B\}.$$

We say that H is *strongly monotonic* iff for all $R^N, Q^N \in \mathscr{D}$, $i \in N$, $a \in H(R^N)$,

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

Danilov [4] demonstrated that strong monotonicity is a necessary and sufficient condition for implementation by Nash equilibrium. Therefore, in order to prove Theorem 4.1, it is sufficient to prove the following lemma.

Lemma 4.4. *Let $H : L^N \to \mathcal{P}_0(A)$. If H is implementable by simple/informed mediated Nash equilibrium then H is strongly monotonic.*

Proof. We shall prove the lemma for simple mediated Nash equilibria; the proof is easily modified for informed mediated Nash equilibria by changing some of the quantifiers. The proof follows the lines of Danilov and Sotskov [5, Theorem 2.3.11].

Let H be a SCC which is implementable by SMNE. Let $R^N \in L^N$, $a \in \pi(\text{SMNE}(\Gamma, R^N))$. We shall show that for all $i \in N$ there exists $\tau_0^{N\setminus\{i\}} \in \Sigma^{N\setminus\{i\}}$ such that for all $\sigma^i \in \Sigma^i$,

$$\pi(\sigma^i, \tau_0^{N\setminus\{i\}}) \in Ess_i(L(a, R^i)). \tag{1}$$

We first claim that this is sufficient to prove the Lemma. Indeed, let $R^N \in L^N$, $a \in H(R^N)$, and $Q^N \in L^N$ such that for all $i \in N$, $Ess_i(L(a, R^i)) \subseteq$

$L(a, Q^i)$; we must show that $a \in H(Q^N)$. Since $a \in H(R^N)$ and Γ implements H, there exists $\sigma_*^N \in \Sigma^N$ such that $\pi(\sigma_*^N) = a$ and, by (1), for all $i \in N$ there exists $\tau_0^{N\setminus\{i\}} \in \Sigma^{N\setminus\{i\}}$ such that for all $\sigma^i \in \Sigma^i$,

$$\pi(\sigma^i, \tau_0^{N\setminus\{i\}}) \in Ess_i(L(a, R^i)) \subseteq L(a, Q^i).$$

This readily implies that σ_*^N is a simple mediated equilibrium point of (Γ, Q^N). As Γ implements H we have that $a = \pi(\sigma_*^N) \in H(Q^N)$.

It remains to prove (1). Let $R^N \in L^N$, $a \in \pi(\text{SMNE}(\Gamma, R^N))$, $i \in N$. Since a is the outcome of a SMNE, it holds that there exists $\tau_0^{N\setminus\{i\}} \in \Sigma^{N\setminus\{i\}}$ such that for all $\sigma^i \in \Sigma^i$,

$$\pi(\sigma^i, \tau_0^{N\setminus\{i\}}) \in L(a, R^i). \tag{2}$$

Let $\sigma^i \in \Sigma^i$, and denote $x = \pi(\sigma^i, \tau_0^{N\setminus\{i\}})$. Let $X = L(a, R^i)$; by (2), $x \in X$.

Now, assume by way of contradiction that $x \notin Ess_i(X)$. Consider the preference profile defined by:

Q^i	$Q^{N\setminus\{i\}}$
$A \setminus X$	x
x	
$X \setminus \{x\}$	$A \setminus \{x\}$

That is, i prefers any alternative in $A \setminus X$ to x, and prefers x to any other alternative in X. The other players prefer x to any other alternative. It holds that $L(x, Q^i) = X$. Therefore, if $x \in H(Q^N)$ we would get that x is essential for i in X, in contradiction to our assumption that $x \notin Ess_i(X)$. We conclude that $x \notin H(Q^N)$; thus, since Γ implements H, we obtain that $x \notin \pi(\text{SMNE}(\Gamma, Q^N))$.

In Q^N, the players in $N \setminus \{i\}$ all rank x first. Hence, the fact that x is not an equilibrium outcome in (Γ, Q^N) can only imply that player i has a strategy which makes it beneficial to deviate, *i.e.* for every $\tau^{N\setminus\{i\}} \in \Sigma^{N\setminus\{i\}}$ there exists $\sigma^i \in \Sigma^i$ such that $\pi(\sigma^i, \tau^{N\setminus\{i\}}) \in A \setminus L(x, Q^i)$. Since $L(x, Q^i) = X = L(a, R^i)$, it follows that for every $\tau^{N\setminus\{i\}} \in \Sigma^{N\setminus\{i\}}$ there exists $\sigma^i \in \Sigma^i$ such that $\pi(\sigma^i, \tau^{N\setminus\{i\}}) \in A \setminus L(a, R^i)$. We have obtained a contradiction to the assumption that $a \in \pi(\text{SMNE}(\Gamma, R^N))$. □

In a sense, Theorem 4.1 implies that in the context of Nash implementation, the presence of mediators cannot aid the social planner. Indeed, by the theorem implementation by Nash equilibrium is easier. Furthermore, implementation by Nash equilibrium is at least as desirable as implementation by mediated Nash, and in many cases (in which a mediator cannot be assumed to be present) strictly more so.

However, it might still be the case that implementation by mediated Nash equilibrium leads to simpler, more natural implementing game forms, when compared with its non-mediated counterpart. We leave this issue as an open question.

5 Implementation by Mediated Strong Equilibrium

We now turn our attention to this paper's main result: a characterization of social choice correspondences implementable by either simple or informed mediated strong equilibria. Subsection 5.1 gives a concise and elegant, but possibly hard to verify, characterization. Subsection 5.2 makes this characterization more tractable by further breaking down the conditions. In Subsection 5.3 we investigate the power of implementation by mediated strong equilibria (using, incidentally, the first characterization).

5.1 First (Concise) Characterization

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

Theorem 5.1. *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 game form 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 game form 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 5.2. *[16, Remarks 6.1.9 and 6.1.15] Let Γ be a game form. Then E^Γ_α and E^Γ_β are monotonic, and E^Γ_α is superadditive.*

Lemma 5.3. *[18, Theorem 4.2] Let $E : \mathcal{P}(N) \to \mathcal{P}(\mathcal{P}_0(A))$ be a stable and maximal effectivity function, 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 $E^{\Gamma^F}_\beta = E$.*

Proof of Theorem 5.1. Assume first that H is implementable by simple mediated strong equilibrium. Let $\Gamma = \langle \Sigma^1, \ldots, \Sigma^n; \pi\rangle$ be the implementing game form; we claim that E^Γ_β is as required.

We first verify that E^Γ_β is indeed an effectivity function. Clearly, for all $S \in \mathcal{P}_0(N)$, $A \in E^\Gamma_\beta(S)$. 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^Γ_β is monotonic with respect to the alternatives by Lemma 5.2) that $E^\Gamma_\beta(N) = \mathcal{P}_0(A)$.

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

$$H(R^N) = \mathrm{SMSE}(\Gamma, R^N) = C(E^\Gamma_\beta, R^N), \tag{3}$$

where the first equality is true as Γ is an implementation of H by SMSE, and the second equality follows from Lemma 3.3.

Finally, we must show that E^{Γ}_{β} is maximal. Let $S \in \mathcal{P}_0(N)$, $B \in \mathcal{P}_0(A)$. If $B \notin E^{\Gamma}_{\beta}$, 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, $A \setminus B \in E^{\Gamma}_{\alpha}(N \setminus S)$, and in particular $A \setminus B \in E^{\Gamma}_{\beta}(N \setminus S)$.

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 5.3, $E^{\Gamma^F}_{\beta} = E$. Therefore, for all $R^N \in L^N$,

$$H^*(R^N) = C(E, R^N) = C(E^{\Gamma^F}_{\beta}, R^N) = \mathrm{SMSE}(\Gamma^F, R^N),$$

where the first equality follows from the assumption, the second by the above-mentioned theorem, and the third equality is a consequence of Lemma 3.3. In particular, for all $R^N \in \mathscr{D}$,

$$H(R^N) = H^*(R^N) = \mathrm{SMSE}(\Gamma^F, R^N).$$

□

Remark 5.4. It is possible to drop the monotonicity of E from the characterization. We leave it in as it provides a unified interface for Theorems 5.1 and 5.5, 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 5.5. *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 5.6. *[17, Theorem 3.5] Let $E : \mathcal{P}(N) \to \mathcal{P}(\mathcal{P}_0(A))$ be an effectivity function. Then E is monotonic and superadditive if, and only if, there exists a game form Γ such that $E = E^{\Gamma}_{\alpha}$.*

Proof of Theorem 5.5. Assume first that H is implementable by informed mediated strong equilibrium. Let Γ be the implementing game form; we will show that E^{Γ}_{α} is as required. As in the proof of Theorem 5.1, E^{Γ}_{α} is an effectivity function due to the attainability assumption. For all $R^N \in \mathscr{D}$ it holds that

$$H(R^N) = \mathrm{IMSE}(\Gamma, R^N) = C(E^{\Gamma}_{\alpha}, R^N). \tag{4}$$

The first equality follows from the fact that Γ is an implementation of H by IMSE; the second equality is implied by Lemma 3.3. By Lemma 5.2, E^{Γ}_{α} 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 5.6, since E is monotonic and superadditive, there exists a game form Γ such that $E = E^{\Gamma}_{\alpha}$; we claim that the foregoing game form Γ implements H by IMSE. Indeed, we have that for all $R^N \in \mathscr{D}$,[1]

$$H(R^N) = C(E, R^N) = C(E^{\Gamma}_{\alpha}, R^N) = \mathrm{IMSE}(\Gamma, R^N),$$

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 3.3. □

5.2 Second (Tractable) Characterization

Although Theorems 5.1 and 5.5 give aesthetic 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 effectivity function with certain properties. The key to simplicity, in this context, is the observation that this effectivity function must be chosen to be $E^*(H)$. Indeed, the following lemma is previously known.

Lemma 5.7. *[16, 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 $E^*(H) = E$.*

We shall now formulate our simplification theorem. We shall obtain more tractable, albeit less concise, characterizations as an easy corollary. In contrast to Subsection 5.1, heretofore the results are formulated for social choice correspondences whose domain is the universal domain L^N. We shall need the following definition [16, 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 5.8. *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 5.9. *[16, 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.*

[1] As in the proof of Theorem 5.1, 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.

Lemma 5.10. *[16, 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 5.11. *[16, 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 5.8. 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 $i \in N$, xP^iy. Since E is an effectivity function, $\{x\} \in E(N)$. 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 5.9. Moreover, By Lemma 5.7, $E = E^*$. 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 5.11, $H(R^N) \subseteq C(E^*, R^N)$ 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 5.10, E^* is monotonic.

We argue that since H is Pareto optimal, E^* is an effectivity function. 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 social choice correspondences which are implementable by mediated equilibria, by combining Theorems 5.1, 5.5, and 5.8.

Corollary 5.12. *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 5.8's conditions 1–3 hold and $E^*(H)$ is maximal.*
2. *H is implementable by informed mediated strong equilibrium if, and only if, Theorem 5.8's conditions 1–3 hold and $E^*(H)$ is superadditive.*

Remark 5.13. Theorem 5.1 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-effectivity function* 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 5.8 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 $E^* = E^*(H)$.

5.3 The Power of Implementation by Strong Mediated Equilibria

In this subsection 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 [6, 7]. Alas, these characterizations are rather hard to formulate, if not to verify. So, one immediate advantage of implementation by mediated strong equilibria is the elegance of Theorems 5.1 and 5.5.

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

Theorem 5.14. *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 5.15. *[16, 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 5.14. By Theorem 5.1, 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 5.15 H is implementable by SE. □

Theorem 5.14 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 5.1 states that the implementing game form may be chosen to be a SCF. This is no small thing; the implementing game form 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 game form be as simple as possible. The implementing game form given in the proof of Lemma 5.15, 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 5.15 is defined directly from the EF E, while the one constructed in the proof of Theorem 5.1 requires the computation of the core of an EF; this task may prove intractable [13].

We move on to implementation by IMSE. As mentioned in Section 4, 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 which is monotonic, superadditive, and stable, but not maximal. By Theorem 5.5, $C(E, \cdot)$ is implementable by IMSE; by Lemma 5.15, $C(E, \cdot)$ is not implementable by SE.

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

Example 5.16 (SCC which 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 3.4, 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 game form). Define an effectivity function 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 5.5, P is implementable by IMSE.

On the other hand, it is also straightforward that E is not maximal, and thus according to Lemma 5.15 P is not implementable by SE.

Example 5.17 (SCC which 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 effectivity function $E_* = E_*(\beta_*)$ 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:

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

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

Peleg [16, 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 effectivity function. By Theorem 5.5, 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, Winter and Peleg [19] prove:

Lemma 5.18. *[19, 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 5.1 and 5.5 we obtain:

Theorem 5.19. *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 first result established that any SCC implementable by SMNE or IMNE is also implementable by NE. We left it as an open question whether, in the case of Nash implementation, mediators can help simplify the implementing game form.

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 game form 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

[1] J. Abdou. Nash and strongly consistent two-player game forms. *International Journal of Game Theory*, 24:345–356, 1995.

[2] J. Abdou and H. Keiding. On necessary and sufficient conditions for solvability of game forms. *Mathematical Social Sciences*, 46(3):243–260, 2003.

[3] 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.

[4] V. I. Danilov. Implementation via Nash equilibria. *Econometrica*, 60(1):43–56, 1992.

[5] V. I. Danilov and A. I. Sotskov. *Social Choice Mechanisms*. Springer, 2002.

[6] B. Dutta and A. Sen. Implementation under strong equilibrium: A complete characterization. *Journal of Mathematical Economics*, 20:49–67, 1991.

[7] P. Fristrup and H. Keiding. Strongly implementable social choice correspondences and the supernucleus. *Social Choice and Welfare*, 18:213–226, 2001.

[8] V. A. Gurvich. Equilibrium in pure strategies. *Soviet Mathematics Doklady*, 38:597–602, 1989.

[9] L. Hurwicz. Optimality and informational efficiency in resource allocation processes. In K. J. Arrow, S. Karlin, and P. Suppes, editors, *Mathematical Methods in the Social Sciences*, pages 27–46. Stanford University Press, 1960.

[10] L. Hurwicz. On informationally decentralized systems. In R. Radner and C. B. McGuire, editors, *Decision and Organization*, pages 297–336. North Holland, 1972.

[11] E. Maskin. Implementation and strong nash equilibrium. In J.J. Laffont, editor, *Aggregation and revelation of preferences*, pages 433–439. North-Holland, 1979.

[12] E. Maskin. Nash equilibrium and welfare optimality. *Review of Economic Studies*, 66:23–38, 1999. This paper was first circulated in 1977.

[13] M.Mizutani, Y. Hiraide, and H. Nishino. Computational complexity to verify the unstability of effectivity function. *International Journal of Game Theory*, 22(3):225–239, 1993.

[14] D. Monderer and M. Tennenholtz. Strong mediated equilibrium. In *Proceedings of the 21st National Conference on Artificial Intelligence*, 2006.

[15] H. Moulin and B. Peleg. Cores of effectivity functions and implementation theory. *Journal of Mathematical Economics*, 10:115–145, 1982.

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

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

[18] B. Peleg and A. D. Procaccia. Mediators enable truthful voting. Discussion paper 451, Center for the Study of Rationality, The Hebrew University of Jerusalem, 2007.

[19] B. Peleg and E. Winter. Constitutional implementation. *Review of Economic Design*, 7:187–204, 2004.

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

A Mediators and Consistency

In this appendix we establish several results concerning game forms which are consistent with respect to mediated equilibria. We say that a GF Γ is NE-consistent (resp. SE-consistent, SMNE-consistent, IMNE-consistent, SMSE-consistent, IMSE-consistent) if for all $R^N \in L^N$, $\mathrm{NE}(\Gamma, R^N) \neq \emptyset$ (respectively $\mathrm{SE}(\Gamma, R^N) \neq \emptyset$, $\mathrm{SMNE}(\Gamma, R^N) \neq \emptyset$, $\mathrm{IMNE}(\Gamma, R^N) \neq \emptyset$, $\mathrm{SMSE}(\Gamma, R^N) \neq \emptyset$, $\mathrm{IMSE}(\Gamma, R^N) \neq \emptyset$). Consistency is a worthy agenda in its own right (see, *e.g.*, [8, 1, 2]), but is also related to implementation. Indeed, if a GF Γ implements a SCC H by some equilibrium concept, then Γ must be consistent with respect

to that selfsame concept. Before proceeding to our results, we note that in the following, the outcome function π of all GFs is assumed to be onto A.

We first notice that Lemma 3.3 directly yields an elegant characterization of GFs which are consistent with respect to strong mediated equilibria.

Corollary A.1. *Let $\Gamma = \langle \Sigma^1, \ldots, \Sigma^n; \pi \rangle$ be a GF. Then:*

1. *Γ is SMSE-consistent if, and only if, E^Γ_β is stable.*
2. *Γ is IMSE-consistent if, and only if, E^Γ_α is stable.*

Abdou and Keiding [2] note that, for a GF Γ, E^Γ_β is stable iff E^Γ_α is maximal and subadditive. This straightforwardly gives us an additional corollary.

Corollary A.2. *Let $\Gamma = \langle \Sigma^1, \ldots, \Sigma^n; \pi \rangle$ be a GF. Then Γ is SMSE-consistent if, and only if, E^Γ_α is maximal and subadditive.*

We next show that, when $n = 2$, NE, SMNE, and SMSE-consistency are equivalent.

Theorem A.3. *Let $\Gamma = \langle \Sigma^1, \Sigma^2; \pi \rangle$ be a two-player GF. Then the following conditions are equivalent:*

1. *Γ is NE-consistent.*
2. *Γ is SMNE-consistent.*
3. *Γ is SMSE-consistent.*

Note that, even though $n = 2$, SMNE-consistency is not trivially equivalent to SMSE-consistency, as a SMSE outcome is required to be Pareto optimal, whereas this is not the case with respect to a SMNE outcome. In order to prove the theorem, we require the following Lemma.

Lemma A.4. *[1, Theorem 2.3] Let $\Gamma = \langle \Sigma^1, \Sigma^2; \pi \rangle$ be a two-player GF. Then the following conditions are equivalent:*

1. *Γ is NE-consistent*
2. *E^Γ_β is stable.*
3. *E^Γ_α is maximal.*

Proof of Theorem A.3. By Corollary A.1 and item 2 in Lemma A.4, it is clear that NE-consistency is equivalent to SMSE-consistency. Also note that as any NE is a SMNE, NE-consistency implies SMNE-consistency. Therefore, it only remains to prove the converse: SMNE-consistency implies NE-consistency; by Lemma A.4, it is sufficient to prove that if Γ is SMNE-consistent, then E^Γ_α is maximal.

Indeed, assume without loss of generality that there exists $B \in \mathcal{P}_0(A)$ such that $B \notin E^\Gamma_\alpha(\{1\})$; we must show that $A \setminus B \in E^\Gamma_\alpha(\{2\})$. Consider the preference profile R^N given by:

R^1	R^2
B	$A \setminus B$
$A \setminus B$	B

Since Γ is SMNE-consistent, there exists $x \in \pi(\text{SMNE}(\Gamma, R^N))$. We claim that $x \notin B$. Indeed, otherwise there exists $\sigma^1 \in \Sigma^1$ such that for all $\sigma^2 \in \Sigma^2$, $\pi(\sigma^N) \in L(x, R^2) \subseteq B$. This is a contradiction to the assumption that $B \notin E^{\Gamma}_{\alpha}(\{1\})$. Hence, $x \in A \setminus B$. That is, there exists $\sigma^2 \in \Sigma^2$ such that for all $\sigma^1 \in \Sigma^1$, $\pi(\sigma^N) \in L(a, R^1) \subseteq A \setminus B$. It follows that $A \setminus B \in E^{\Gamma}_{\alpha}(\{2\})$. $\square$

Is it also true that NE-consistency is equivalent to SMNE-consistency when the number of players is greater than two? This remains an interesting open question.

Example A.5 (IMSE-consistent GF which is not SMSE-consistent)**.** Consider the Modulo Game Γ, given in Example 3.4; by that example, Γ is IMSE-consistent. However, it is also easy to see that if $R^N \in L^N$ is not a unanimous profile, *i.e.* there is no single alternative which all players rank first, then (Γ, R^N) has no SMSE (and in particular, no SE).

Example A.6 (SMSE-consistent GF which is not SE-consistent)**.** Consider the following two-player GF Γ.

	L	M	R
U	a	a	a
M	a	b	c
D	a	c	b

Clearly Γ is NE-consistent, as $\langle U, L\rangle$ is always a Nash equilibrium. Therefore, by Theorem A.3, Γ is also SMSE-consistent.

Now, consider the preference profile R^N given by:

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

It is easy to verify that (Γ, R^N) has no SE.