Strong Mediated Equilibrium*

Dov Monderer and Moshe Tennenholtz
Faculty of Industrial Engineering and Management
Technion–Israel Institute of Technology
Haifa 32000, Israel

October 26, 2006

Abstract

Providing agents with strategies that will be robust against deviations by coalitions is central to the design of multi-agent agents. However, such strategies, captured by the notion of strong equilibrium, rarely exist. This paper suggests the use of mediators in order to enrich the set of situations where we can obtain stability against deviations by coalitions. A mediator is a reliable entity, which can ask the agents for the right to play on their behalf, and is guaranteed to behave in a pre-specified way based on messages received from the agents. However, a mediator can not enforce behavior; that is, agents can play in the game directly without the mediator's help. We prove some general results about mediators, and concentrate on the notion of strong mediated equilibrium; we show that desired behaviors, which are

*This version is a slightly modified version of the manuscript appearing at the proceedings of AAAI 2006. However, many of the proofs are not included yet.

stable against deviations by coalitions, can be obtained using mediators in a rich class of settings.

1 Introduction

In the recent years there has been much interest in work bridging AI and game theory. This interest is a result of the need to consider agents' incentives in multi-agent systems. Indeed, when considering a prescribed multi-agent behavior, it makes little sense to assume that an agent will stick to its part of that behavior, if deviating from it can increase its payoff. This leads to much interest in the study of equilibrium in games; a Nash equilibrium is a strategy profile for the agents, such that deviation by each single agent is not beneficial for it. The good news is that when agents are allowed to use mixed strategies then a Nash equilibrium always exists; the bad news however is that this concept does not take into account deviations by non-singleton sets of agents. While stability against deviations by subsets of the agents is a most natural requirement, it is well known that obtaining such stability is possible only in extremely rare situations.

In order to tackle this issue we consider in this paper the use of *mediators*. A mediator is a reliable entity that can interact with the players and perform on their behalf actions in a given game. However, a mediator can not enforce behavior. Indeed, an agent is free to participate in the game without the help of the mediator. This notion is highly natural; in many systems there is some form of reliable party or administrator that can be used as a mediator. Notice that we assume that the multi-agent interaction (formalized as a game) is given, and all the mediator can do is to communicate with the agents and perform actions on behalf of the agents that allow it to do so. The mediator's behavior is pre-specified and depends on the messages received from all agents. This natural setting is different from the one

discussed in the theory of mechanism design (see Jackson (2001) for an introduction) where a designer designs a new game from scratch in order to yield some desired behavior. The simplest form of mediator discussed in the game theory literature is captured by the notion of correlated equilibrium (Aumann (1974)). This notion was generalized to communication equilibrium by Forges (1986); Myerson (1986). Another, more powerful type of mediators is discussed in Monderer & Tennenholtz (2004). However, in all these settings the mediator can not perform actions on behalf of the agents that allow it to do so.

We find the notion of a mediator as central to the theory of multi-agent systems. As a result, in this paper we develop a rigorous concept of a mediator, and prove its generality. We concentrate on the notion of strong mediated equilibrium, a multi-agent behavior which is robust against deviations by coalitions in a mediated game.

In order to illustrate the power of a reliable mediator as discussed in this paper, consider the famous prisoners dilemma game:

	Cooperate	Defect
Cooperate	4,4	0,6
Defect	6,0	1,1

In this classical example we get that in the only equilibrium both agents will defect, yielding both of them a payoff of 1. However, this equilibrium, which is also a dominant strategy equilibrium, is inefficient; indeed, if both agents deviate from defection to cooperation then both of them will improve their payoffs. Formally, mutual defection is not a strong equilibrium; it is not stable against deviations by coalitions. Indeed, there is no strong equilibrium in the prisoners

dilemma game.

Consider now a reliable mediator who offers the agents the following protocol: if **both** agents agree to use the mediator services then he will perform cooperate on behalf of both agents. However, if only one agent agrees to use his services then he will perform defect on behalf of that agent. Notice that when accepting the mediator's offer the agent is committed to actual behavior as determined by the above protocol. However, there is no way to enforce the agents to accept the suggested protocol, and each agent is free to cooperate or defect without using the mediator's services. Hence, the mediator's protocol generates a new game, which we call *mediated game*:

	Mediator	Cooperate	Defect
Mediator	4,4	6,0	1,1
Cooperate	0,6	4,4	0,6
Defect	1,1	6,0	1,1

The mediated game has a most desirable property: in this game there is a strong equilibrium; that is, equilibrium which is stable against deviations by coalitions. In this equilibrium both agents will use the mediator services, which will lead them to a payoff of 4 each! We call a strong equilibrium in a mediated game: a *strong mediated equilibrium*. Hence, we get cooperation in the prisoners dilemma game as an outcome of a strong mediated equilibrium.[1]

[1]Bernheim, Peleg, & Whinston (1987) defined and analyzed coalition-proof equilibrium, which is a generalization of strong equilibrium for games in strategic form. This type of equilibrium has been further analyzed in e.g.,Einy & Peleg (1995); Moreno & Wooders (1996). This

We put the concept of a mediator in the perspective of work in game theory. In particular, the concept of c-acceptable strategies introduced in Aumann (1959) is an abstract notion that captures the "reasonable outcomes" obtained when subsets of the set of agents can correlate their activities. Our work introduces an explicit model of the mediation activity. We show that the introduction of an explicit mediator makes a difference; the outcomes that can be obtained using strong mediated equilibria are different from the ones that can be obtained using c-acceptable strategies.

Given the general concept of a mediator, we prove that mediators can indeed significantly increase the set of desired outcomes that can be obtained by multi-agent behaviors which are robust against deviations by coalitions. Namely, we show that any symmetric game possesses a strong mediated equilibrium, which also leads to optimal surplus, iff its standard TU core is non-empty. This class includes many natural sub-classes of games. We also show another positive result with regard to deviations of coalitions of size at most k. We show that in any symmetric game with n agents, if $k!$ divides n then there exists a k-strong mediated equilibrium, leading to optimal surplus. However, if $k!$ does not divide n, then one can find a symmetric game (with n agents) where the above property does not hold.

In the last section we discuss the connection between mediators and the notion of program equilibrium (Tennenholtz (2004)). Most proofs are omitted from this paper due to lack of space.

---

enables one to naturally define "coalition-proof mediated equilibrium". We do not analyze this interesting topic in this paper.

2 Games in strategic form: Strong Equilibrium

Some notational preliminaries are needed. When we discuss subsets of a given set we implicitly assume non-emptiness unless we specify otherwise. Let Y be a finite set. The set of probability distributions over Y is denoted by $\Delta(Y)$. That is, every $c \in \Delta(Y)$ is a function $c : Y \to [0,1]$ such that $\sum_{y \in Y} c(y) = 1$. For every $y \in Y$ we denote by δ_y the probability distribution that assigns probability 1 to y. Let I be a finite set of indices, and let $c_i \in \Delta(Y_i)$, $i \in I$, where Y_i is a finite set for every $i \in I$. We denote by $\times_{i \in I} c_i$ the product probability distribution on $Y = \times_{i \in I} Y_i$ that assigns to every $y \in Y$ the probability $\prod_{i \in I} c_i(y_i)$.

A game in strategic form is a tuple $\Gamma = \langle N, (X_i)_{i \in N}, (u_i)_{i \in N} \rangle$, where N is a finite set of players, X_i is the strategy set of player i, and $u_i : X \to \Re$ is the payoff function for player i, where $X = \times_{i \in N} X_i$. If $|N| = n$, whenever convenient we assume $N = \{1, \cdots, n\}$. Γ is *finite* if the strategy sets are finite.

Let $\Gamma = \langle N, (X_i)_{i \in N}, (u_i)_{i \in N} \rangle$ be a finite game. For every $S \subseteq N$ we denote $X_S = \times_{i \in S} X_i$. When the set, N is clear, $X_{N \setminus S}$ will be also denoted by X_{-S}, and moreover, $X_{-\{i\}}$ will be also denoted by X_{-i}. Let $S \subseteq N$, every $c \in \Delta(X_S)$ is called a *correlated strategy for* S . A correlated strategy for the set of all players N is also called a *correlated strategy*, and for every i, a correlated strategy for $\{i\}$ is also called a *mixed strategy* for i. The expected payoff of i with respect to a correlated strategy c is denoted by $U_i(c)$. That is,

$$U_i(c) = \sum_{x \in X} u_i(x) c(x)$$

.

For every S, the set of mixed-strategy profiles is denoted by Q_S. That is, $Q_S = \times_{i \in S} \Delta(X_i)$. We will use Q for Q_N, and Q_i for $Q_{\{i\}}$.

The *mixed extension* of the game Γ is the game $(N, (Q_i)_{i \in N}, (w_i)_{i \in N})$, where for every $q \in Q$, $w_i(q) = U_i(q_1 \times \cdots \times q_n)$.

Strong Equilibrium

Let $\Gamma = \langle N, (X_i)_{i\in N}, (u_i)_{i\in N}\rangle$ be a game in strategic form, and let $x \in X$. We say that x is a *strong equilibrium of type I* in Γ if the following holds:

For every subset, S of players and for every $y_S \in X_S$ there exists $i \in S$ such that $u_i(y_S, x_{-s}) \leq u_i(x)$. When Γ is finite we define two other notions of strong equilibrium: Let $q = (q_1, \cdots, q_n)$ be a profile of mixed strategies.

We say that q is a *strong equilibrium of type II in* Γ if q is a strong equilibrium of type I in the mixed extension of Γ. That is, q is a strong equilibrium of type II in Γ if for every subset of players S, and for every profile of mixed strategies $(p_i)_{i\in S}$ there exists $i \in S$ such that $U_i(\times_{i\in S} p_i \times_{i\in N\setminus S} q_i) \leq U_i(\times_{i\in N} q_i)$. Obviously, if $x \in X$, and $(\delta_{x_1}, \cdots, \delta_{x_n})$ is a strong equilibrium of type II in Γ then x is a strong equilibrium of type I, but the converse does not hold. Whenever $(\delta_{x_1}, \cdots, \delta_{x_n})$ is a strong equilibrium of type II we abuse notations and we allow ourselves to say that x is a strong equilibrium of type II in Γ.

Let $q \in Q$. We say that q is a *strong equilibrium of type III*, if for every subset of players S , and for every correlated strategy for S, $c_S \in \Delta(X_S)$ there exists $i \in S$ such that $U_i(c_S \times (\times_{i\in S^c} q_i)) \leq U_i(q_1 \times q_2 \times \cdots \times q_n)$.

Let $q \in Q$. Obviously q is a strong equilibrium of type II if q is a strong equilibrium of type III, but not vice versa.

Let $q \in Q$. The requirement that q is a strong equilibrium of type II seems to be acceptable in an environment in which the players believe that they and others could not possibly correlate their behavior (e.g., when every player is sitting in a separate room, and there is no communication between the players). However, every player can perform a private randomization. In an environment in which the players do not correlate their strategies, but they may fear/hope that such a correlation is possible, we expect q to be a strong equilibrium of type III in order to be believed/played by the players.

3 Strong Mediated Equilibrium

We now introduce mediators, a general tool for coordinating and influencing agents' behavior in games. A mediator is always assumed to be reliable. However, mediators are classified according to their abilities to interfere in the game.

In this paper we endow the mediator with the ability to play for the players who give him the right to play for them. However, the mediator cannot enforce the players to use his services.

Let Γ be a finite game in strategic form. A *mediator for* Γ is a tuple $((M_i)_{i\in N}, (\mathbf{c}_S)_{S\subseteq N})$, where each M_i is a finite set, for every $S \subseteq N$,$\mathbf{c}_S : M_S \to \Delta(X_S)$,and $M_i \cap X_i = \emptyset$ for every player i.[2]

Every mediator $\mathcal{M}$ for Γ defines a finite game in strategic form, which we call the *mediated game* and denote by $\Gamma(M)$. In the mediated game, the strategy set of player i is $Z_i = X_i \cup M_i$, and the payoff function of i is defined for every $z \in Z$ as follows:

$$u_i^{\mathcal{M}}(z) = U_i(\mathbf{c}_{T_z}(z_{T_z}) \times (\times_{j\in N\setminus T_z}\delta_{z_j})),$$

where $T_z = \{j \in N | z_j \in M_j\}$. That is, T_z is the set of players who use the service of the mediator.

Every correlated strategy in $\Gamma(M)$, $\xi \in \Delta(Z)$ induces a correlated strategy c_ξ in Γ: for every $x \in X$ we have that

$$c_\xi(x) = \sum_{S\subseteq N} \sum_{m_{-S}\in M_{-s}} \xi(x_S, m_{-S})\mathbf{c}_{-S}(m_{-S})(x_{-S})$$

Hence, $U_i^{\mathcal{M}}(\xi) = U_i(c_\xi)$.

Definition: Let Γ be a game in strategic form. A correlated strategy $c \in \Delta(X)$ is a *strong mediated equilibrium* if there exists a mediator for Γ, $\mathcal{M}$, and a vector

[2]The requirement that the intersection of these sets is empty is done only for technical convenience.

of messages $m \in M = \times_{i \in N} M_i$, with $\mathbf{c}_N(m) = c$, such that m is a strong equilibrium of type III in $\Gamma(\mathcal{M})$. Such a mediator is said to *implement* c.

First, it is immediate to see that by using mediators we don't lose any outcome that can be obtained in a strong equilibrium of the original game:

Proposition 1 *Let Γ be a finite game in strategic form, and let q be a profile of mixed strategies, which is a strong equilibrium of type III in Γ. Then, $q_1 \times q2 \times \cdots \times q_n$ is a strong mediated equilibrium in Γ.*

As we will see, mediators allow to significantly expand in rich classes of games the set of outcomes that are implemented in a way which is robust against deviations by coalitions.

Hence, we focus on Mediators that generate a game, and a particular type of pure strategy strong equilibrium of type III in this game. However, the game generated by the mediator may give rise to other possibilities of forming a strong equilibrium of type III. It is therefore important to know that our seemingly restricted definition does not restrict the possible acceptable outcomes. Indeed, we show:

Proposition 2 *Let $\mathcal{M}$ be a mediator for Γ, and let $\hat{q}$ be a vector of mixed strategies in $\Gamma(\mathcal{M})$, which is a strong equilibrium of type III in $\Gamma(\mathcal{M})$. Then $c_{\hat{q}_1 \times \hat{q}_2 \times \cdots \times \hat{q}_n}$ is a strong mediated equilibrium in Γ.*

Furthermore, when there is one mediator, other mediators may show up. Any set of mediators $\hat{\mathcal{M}}$ generates a game in strategic form $\Gamma(\hat{\mathcal{M}})$ in which every player can choose any mediator in $\hat{\mathcal{M}}$ she wishes, and give this mediator the right to play by sending him a message ,or play independently. If ξ is a correlated strategy in $\Gamma(\hat{\mathcal{M}})$ we denote by c_ξ the correlated strategy in Γ generated by ξ.

Proposition 3 *Let $\hat{\mathcal{M}}$ be a set of mediators, and let $\bar{q}$ be a strong equilibrium of type III in $\Gamma(\hat{\mathcal{M}})$, then $c_{\bar{q}_1 \times \bar{q}_2 \times \cdots \times \bar{q}_b}$ is a strong mediated equilibrium in Γ.*

The above proposition is quite general and shows that the existence of many mediators that the agents can approach does not help beyond the use of our single mediator. Finally we would like to show that a new mediator for the game generated by a mediator cannot add acceptable outcomes:

Proposition 4 *Let Γ be a game in strategic form, and let $\mathcal{M}$ be a mediator for Γ, if ξ is a strong mediated equilibrium in $\Gamma(\mathcal{M})$, then c_ξ is a strong mediated equilibrium in Γ.*

The above results prove the generality of our setup. We now define a type of minimal mediators that will play an important role in our subsequent analysis. Let Γ be a game in strategic form. A mediator $\mathcal{M}$ is minimal if each message space is a singleton. Consider a minimal mediator in which $M_i = \{r_i\}$ for every player i. Let $r = (r_1, \cdots, r_n)$. When the players are using this mediator each of them can either gives the right to play to the mediator (sending r_i) or play independently. If the set of players that give the right to play is T, the mediator use the correlated strategy $c_T = \mathbf{c}_T(r_T)$ in order to play for T. Hence, every minimal mediator is uniquely defined by a vector of correlated strategies, $(c_S)_{S \subseteq N}$, one for each subset of players. As it turns out, restricting our attention to minimal mediators does not cause any loss of strong mediated equilibria:

Lemma 1 *Let Γ be a finite game in strategic form. Every strong mediated equilibrium in Γ can be implemented by a minimal mediator.*

Proof: Let $c \in \Delta(X)$ be a strong mediated equilibrium, and let $\mathcal{M} = ((M_i)_{i\in N}, (\mathbf{c}_S)_{S\subseteq N})$ implement c, by the profile $m \in M_N$. That is, m is a strong equilibrium of type III in $\Gamma(\mathcal{M})$. Define a minimal mediator in which the set of messages for every i is $K_i = \{m_i\}$. And the implementing functions are for every S the restriction of $\mathbf{C}_S$ to $\{m_S\}$. Hence, $\Gamma(\mathcal{K})$ is obtained from $\Gamma(\mathcal{M})$ by restricting the strategy set of every player. Therefore m remains a strong equilibrium of type III in $\Gamma(\mathcal{K})$. ∎

Given the general setup above, it is important to see how it fits previous foundational work in game theory. Aumann (see Aumann (1959)) defined **c-acceptable correlated strategies**, which may seem at first glance to implicitly catch the idea of defining the "reasonable outcomes" that can be obtained when agents can correlated their strategies. A correlated strategy c is c-acceptable if there exists a vector $(c_S)_{S\subseteq N}$ with $c_N = c$ such that for every subset S and for every $d_S \in \Delta(X_S)$, there exists $i \in S$ such that

$$U_i(d_S \times c_{-S}) \leq U_i(c)$$

.

It is easy to show:

Proposition 5 *Every strong mediated equilibrium is c-acceptable.*

Proof: Let c be a strong mediator equilibrium. Therefore there exists a minimal mediator that implements c. Assume this mediator is defined by $(c_S)_{S\subseteq N}$, then obviously this vector satisfies the conditions for c-acceptability: For every S and for every $\xi_S \in \Delta(Z_S)$ there exists $i \in S$, which gets at most $U_i(c)$. As $X_S \subset Z_S$, $\Delta(X_S)$ can be naturally identified with the subset of $\Delta(Z_S)$ of all correlated strategies ξ_S that satisfy $\xi_S(M_S) = 0$. It is therefore obvious that the players in S cannot get better off by deviating to $d_S \in \Delta(X_s)$. ∎

However, as we will show the converse of the above result is not true, for illuminating reasons. From the proof one should notice that all that we used was the

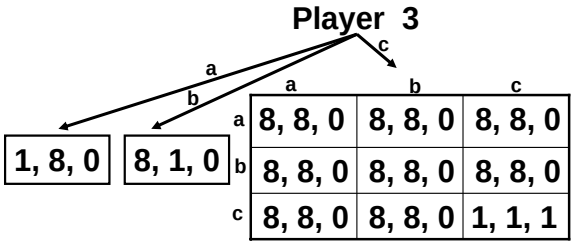
fact that a strong mediated equilibrium is immune from deviation of correlating strategies in the original game. But the concept of mediated equilibrium requires also that it is immune to correlation over messages and actions. Hence, the converse to Proposition 5 is not clear. Indeed we are about to prove that the converse is not true. A c-acceptable strategy used by a mediator is not necessary immune to *Trojan horses*. A subset may correlate in such a way that in some realizations a subgroup is pretending to cooperate by sending the right messages. This may be beneficial because these trojan horses will be part of the punishing group, and they may get a very big payoff for this!

We now show:

Theorem 1 *There exists a game* Γ*, and a c-acceptable strategy* c*, which is not a strong mediated equilibrium.*

Proof:

Consider the following 3-person game:



The strategy set of each player i is $\{a, b, c\}$. If Player 3 chooses a the resulting payoff vector is $(1, 8, 0)$ independent of the actions chosen by 1 and 2. Similarly, when 3 chooses b, the resulting payoff is $(8, 1, 0)$. When 3 chooses c, the resulting payoff matrix is the 3×3 matrix in the figure.

We first show that the correlated strategy that is concentrated on (c, c, c) is c-acceptable. We define a vector $(\eta_S)_{S\subseteq\{1,2,3\}}$, where $\eta_{\{1,2,3\}} = \delta_{(c,c,c)}$ that satisfies the conditions for $\eta_{\{1,2,3\}}$ to be c-acceptable.

Indeed, let $\eta_{\{1,2\}} = (a, a)$, $\eta_{\{1,3\}} = (a, b)$, $\eta_{\{2,3\}} = (a, a)$,$\eta_3 = a$, $\eta_1 = a$,$\eta_2 = a$. Obviously, these are punishing strategies that do the job. No deviating group can ensure more than 1 to each of its members. More generally, such punishment will be obtained when the following are satisfied: $\eta_1, \eta_2, \eta_{\{1,2\}}$ can be anything, $\eta_{\{1,3\}}$ assigns probability 1 of doing b by 3, $\eta_{\{2,3\}}$ assigns probability 1 of doing a by 3, and η_3 assigns probability 1 to doing a by 3, or probability 1 to doing b by 3. Notice that in defining η_3 3 can punish one of the players 1 and 2, that is η_3 must be a or b, but 3 cannot randomize between these two punishments.

Hence, if (c, c, c) is a a strong mediated equilibrium, and $\mathcal{M}$ is a minimal mediator that implements it, we can assume without lose of generality that whenever the set of players that give the right to play to the mediator is T, the mediator play η_T. However, in $\Gamma(\mathcal{M})$, 1 and 2 can randomize with equal probability between the two options: "1 plays a and 2 give the right to play to the mediator", and "2 plays a and 1 give the right to play to the mediator". This will give each of them an expected payoff of 4.5. Hence, (c, c, c) is not a strong mediated equilibrium. ∎

The above result shows that strong-mediated equilibria are different from c-acceptable strategies. It is not just that our study builds on explicit model of general mediators, it also shows a limitation of c-acceptable strategies as "reasonable outcomes" in situations where agents can correlated their activities.

For completeness, we add now an analysis of strong mediated equilibria in the game in the proof of Theorem 1. Note that the correlated strategy, c which randomizes with equal probability between (c, c, c) and (a, a, c) is a strong mediated equilibrium in the above game. In this strong mediated equilibrium, 1, and 2 get each 4.5, and 3 gets 0.5. Note also that $(4.5.4.5, 0.5)$ is the unique payoff vector that can be obtained in a strong mediated equilibrium in this game.

3.1 The β-core and m-core

The vector of payoffs obtained using c-acceptable correlated strategies are called the β-core (see Aumann (1961)). We now recall this notion, and introduce the m-core which refers to the notion of strong mediates equilibrium. This will provide an interpretation of the previous result from the point of view of the payoff vectors that can be obtained. Let Γ be a game in strategic form. A vector of payoffs $\mathbf{w} = (w_1, \cdots, w_n)$ is *a strong mediated payoff vector,* $\mathbf{w} \in \mathbf{C}_m(\Gamma)$ (resp. in the β-core of Γ, $\mathbf{w} \in \mathbf{C}_\beta(\Gamma)$) if there exists a a strong mediated equilibrium (resp. c-acceptable) correlated strategy c such that $U_i(c) = w_i$ for every player i. We define the m-core as the corresponding set of strong mediated payoff vectors. Hence, our previous results imply that the β-core of a game contains its m-core, and there are games in which the β-core strictly contains the m-core.

4 Existence

In this section we show the power of mediators, by showing rich settings where desired outcomes can be implemented using mediators in a way which is stable against deviations by coalitions.

In Aumann (1959) Aumann proved that every 2-person game has a c-acceptable

strategy.

Proposition 6 *In a two-person game, every c-acceptable strategy is a strong mediated equilibrium. Consequently, Every 2-person game has a strong mediated equilibrium.*

Before presenting our other existence results we need to define the notion of symmetry.

A *permutation* of the set of players is a one-to-one function from N onto N. For every permutation π, and for every action profile $x \in X$ we denote by πx the permutation of x by π. That is, $(\pi x)_{\pi i} = x_i$ for every player $i \in S$. Γ is a *symmetric game* if $X_i = X_j$ for all $i, j \in N$, and $u_i(x) = u_{\pi(i)}(\pi x)$ for every player i, for every action profile $x \in X$, and for every permutation π.

Our idea will be to consider symmetric games. Needless to say that symmetric games are most popular in computerized settings. For example, the extremely rich literature on congestion games in computer science deals with particular form of symmetric games.

In order to show a general existence result for strong mediated equilibrium in the context of symmetric games, we will use the standard game-theoretic notions on cooperative games, and in particular the Transferable Utility (TU) Core. For an introduction to this subject the reader may consult Aumann & Hart (1992).

Let Γ be a game in strategic form. For every $x \in X$ and for every $S \subseteq N$ we let $u_S(x) = \sum_{i \in S} u_i(x)$. Similarly, for a correlated strategy c we denote $U_S(c) = \sum_{i \in S} U_i(c)$. For every S let

$$v(S) = \min_{c_{-s}} \max_{c_S} U_S(c_S, c_{-S}).$$

By the minimax theorem,

$$v(S) = \max_{c_S} \min_{c_{-s}} U_S(c_S, c_{-S}).$$

A strategy c_S is an optimal strategy for S if it guarantees $v(S)$ to the members of S. That is, the max is attained in c_S. A strategy c_S is an optimal punishing strategy for S if it ensures that N/S does not obtain more than $v(N/S)$. v is a TU-cooperative game. A payoff vector $(w_1, w_2, \ldots, w_n)$ is in the core, $C(v)$ if

- $\sum_{i=1}^{n} w_i = v(N)$;
- $\sum_{i \in S} w_i \geq v(S)$ for every $S \subset N$.

Obviously, if Γ is a symmetric game, v is a symmetric TU-game, that is $v(\pi S) = v(S)$ for every $S \subseteq N$ and every permutation π. It is obvious that the core of a symmetric TU game is symmetric, and therefore it must contain a symmetric payoff vector w, that is $w_i = \frac{v(N)}{n}$ for every i.

The TU-core, or core for short, is not always non-empty, but it is much larger set of payoffs than the one that can be obtained in a strong equilibrium of type III in a game in strategic form. Fortunately, we can show the following:

Theorem 2 *Let Γ be a symmetric game. Let v be its associated TU-game. If $C(v) \neq \emptyset$, then there exists a strong mediated equilibrium with $U_i(c) = \frac{v(N)}{n}$ for every i. Moreover, there exists a symmetric game with $C(v) = \emptyset$ for which a strong mediated equilibrium does not exist.*

Hence, our results show that having a non-empty (standard) core is in a sense a necessary and sufficient condition for the existence of strong mediated equilibrium. In the related strong mediated equilibrium the agents get optimal surplus (sum of agents' payoffs is maximal) This result points to the high power of mediators in obtaining stability against group deviations, as well as characterize the power of mediators in symmetric games. An example of a natural set of symmetric games with non-empty core is Congestion Games with Failures (CGFs)(

Penn, Polukarov, & Tennenholtz (2005)). In a CGF there are n agents and m resources. Each resource is associated with an increasing delay function, and with a non-negative probability of failure. Each agent has a task, and a worth of v for successful execution of that task. Each agent can select any subset of the resources in order to try and execute its task. Hence, the expected payoff of an agent is v times the probability of success of completing the task by at least one of the chosen resources, minus the minimal delay in executing the task (by at least one of the successful resources). It is easy to see that strong equilibrium (of any type) can not be guaranteed for this class of games; however, by Theorem 2 a strong mediated equilibrium exists in any game of that class.

4.1 k-Strong Mediated Equilibrium

For every "strong" equilibrium concept, and for every $1 \leq k \leq n$ one can define the analog concept of a k-strong equilibrium, in which it is only required that deviation of every subset with at most k players is not profitable. Obviously a 1-strong equilibrium concept is just a Nash equilibrium concept, and a n-strong equilibrium concept is simply the corresponding strong equilibrium concept. The notion of k-strong equilibrium is very natural; it captures the idea that it makes sense to coordinate deviation by a set of agents, but this coordination makes sense only of that group is of limited size.

We can prove the following general result:

Theorem 3 *Let Γ be a symmetric game in strategic form. Let $1 \leq k \leq n$ be an integer. If $k!$ divides n there exists a symmetric k-strong mediator equilibrium, leading to optimal surplus. Moreover, if $k!$ does not divide n there exists a symmetric game γ with n players in which there does not exist a k-strong mediated equilibrium.*

Notice that this result implies for example that in any game with even number of agents, a 2-strong mediated equilibrium always exists. The proof shows that in particular there is such an equilibrium where the agents' sum of payoffs (aka their social surplus) is maximized. Hence, optimal social surplus can be obtained using a mediator where deviation by pairs of players are not beneficial!

5 Program equilibrium: a special type of mediators

The theory of mediators discussed in this paper is a very broad one. Indeed, in general, the agents' messages can be arbitrary, and the interpretation of these message can be arbitrary. One interesting type of messages are those that have the flavor of a standard computer program, using a standard programming language; in this case it is interesting to look at mediators whose role is the mere execution of the programs. As shown in Tennenholtz (2004) this perspective can be highly productive. We now briefly discuss program equilibrium and its relationships to our setting.

Consider the prisoners dilemma, discussed in the introduction. Denote the possible actions by D (*defect*) and C (*cooperate*). Recall that in the only (dominant strategy) equilibrium of this game both agents will choose D, while mutual cooperation will lead both of the agents to a higher payoff. In Tennenholtz (2004) this issue has been addressed by considering agents who can use computer programs as their strategies; these computer programs are to run on a single server/machine, and therefore can exploit the famous dual role of computer programs (introduced in von Neumann (1945)): a program can serve both as a set of instructions and as a data file. Consider the program: *IF MY-PROGRAM=YOUR-PROGRAM then C; else D;* The exact syntax and semantics of such programs is discussed in Tennenholtz (2004), but the reader can easily notice the basic idea:

the agent/programmer *instructs* the computer to compare its program to the other program, *as files,* and execute a particular action based on the result of that comparison. There are no circular arguments here, due to the dual role of computer programs. Moreover, this program defines a *program equilibrium*: it is irrational for an agent to deviate from that program assuming the other agent stick to it. As a result, we get cooperation in the one-shot prisoners' dilemma! This result is then extended to a general folk theorem.

An interesting question is whether the role of a mediator can be replaced by a computer program as discussed in Tennenholtz (2004). We can show:

Theorem 4 *Given a game in strategic form* Γ. *Then* Γ *possesses a program equilibrium iff* Γ *possesses a 1-strong (Nash) mediated equilibrium.*

References

Aumann, R. J., and Hart, S. 1992. *Handbook of Game Theory with Economic Applications*. Elsevier Science.

Aumann, R. 1959. Acceptable points in general cooperative n-person games. In Tucker, A., and Luce, R., eds., *Contribution to the Thoery of Games, Vol. IV, Annals of Mathematics Studies, 40*. 287–324.

Aumann, R. 1961. The core of cooperative games without payments. Transactions of the American Mathematical Society.

Aumann, R. 1974. Subjectivity and correlation in randomized strategies. *Journal of Mathematical Economics* 1:67–96.

Bernheim, B. D.; Peleg, B.; and Whinston, M. 1987. Coalition proof Nash equilibrium: I concepts. *Journal of Economic Theory* 42(1):1–12.

Einy, E., and Peleg, B. 1995. *Coalition-Proof Communication Equilibria*. Social Choice, Welfare, and Ethics. Cambridge: Cambridge Univ. Press. chapter 13.

Forges, F. M. 1986. An approach to communication equilibria. *Econometrica* 54(6):1375–85.

Jackson, M. 2001. A Crash Course in Implementation Theory. *Social Choice and Welfare* 18(4):655–708.

Monderer, D., and Tennenholtz, M. 2004. K-Implementation. *Journal of Artificial Intelligence Research (JAIR)* 21:37–62.

Moreno, D., and Wooders, J. 1996. Coalition-proof equilibrium. *Games and Economic Behavior* 17(1):80–112.

Myerson, R. B. 1986. Multistage games with communication. *Econometrica* 54(2):323–358.

Penn, M.; Polukarov, M.; and Tennenholtz, M. 2005. Congestion games with failures. In *EC-05, ACM Conference on Electronic Commerce*, 259–268.

Tennenholtz, M. 2004. Program equilibrium. *Games and Economic Behavior* 49:363–373.

von Neumann, J. 1945. First draft of a report on the edvac, contract no. w-670-ord-402 moore school of electrical engineering, univ. of penn., philadelphia. Reprinted (in part) in Randell, Brian. 1982. Origins of Digital Computers: Selected Papers, Springer-Verlag, Berlin Heidelberg, pp. 383–392.