Minority-proof cheap-talk protocol

Heller, Yuval

Tel-Aviv University

19 August 2005

Minority-Proof Cheap-Talk Protocol

Yuval Heller[1]

September 2009

(First version received in August 2005)

School of Mathematical Sciences,

Tel-Aviv University,

Tel-Aviv 69978, Israel.

Email: helleryu@post.tau.ac.il

Phone: 972-3-6405386, 972-52-5282182

Fax: 972-3-640-9357

[1] The School of Mathematical Sciences, Tel Aviv University. This paper is based on a Master thesis the author done under the supervision of Prof. Ehud Lehrer. I would like to thank Prof. Lehrer for his careful supervision and for the continuous help he offered. My deep gratitude is also given to Prof. Eilon Solan for many discussions and useful ideas concerning the subject, and to the associate editor and the anonymous referees for many useful comments during the process of writing this paper.

Abstract

This paper analyzes the implementation of correlated equilibria that are immune to joint deviations of coalitions by cheap-talk protocols. We construct a cheap-talk protocol that is resistant to deviations of fewer than half the players, and using it, we show that a large set of correlated equilibria can be implemented as Nash equilibria in the extended game with cheap-talk. Furthermore, we demonstrate that in general there is no cheap-talk protocol that is resistant for deviations of half the players.

JEL classification: C72

Keywords: non-cooperative games, cheap-talk, correlated equilibrium, strong equilibrium, coalition-proof equilibrium, fault-tolerant distributed computation.

1. Introduction

Mediated communication allows the implementation of correlated equilibria (Aumann, 1974). In many environments, it is hard to find a fair mediator, and it is natural to ask what can be implemented using only cheap-talk: pre-play, unmediated, non-binding, and non-verifiable communication among the players (See e.g., Crawford & Sobel, 1982; Forges, 1990; Barany, 1992). Coalitions can use cheap-talk to coordinate joint deviations. Aumann (1959) discusses such deviations, and defines a strong Nash equilibrium as a strategy profile from which no coalitional deviation is profitable for all the deviators. Moreno & Wooders (1996) give the correlated counterpart definition: a strong correlated equilibrium.[2]

In some real-world environments, while it is easy for small coalitions to deviate from equilibrium, it is much harder for large coalitions to deviate while hiding it from the non-deviating players.[3] One example for such environment is the field of foreign affairs: there are no known examples of secret joint deviations of large coalitions (a few dozen countries), but there are secret joint deviations of small coalitions (a few countries), as in the following examples:

- The secret additional protocol of Molotov-Ribbentrop Pact (1939) in which two countries have secretly divided between them six neighboring countries.
- The surprising joint attack of Egypt and Syria against Israel in 1973 October war.

We introduce two new concepts:[4] *k-strong correlated equilibrium* and *k-strong Nash equilibrium*, which require resistance against coalitional deviations of up to k players. This paper deals with a cheap-talk protocol that implements a k-strong correlated equilibrium, as a k-strong Nash equilibrium of an extended game. This protocol generalizes existing protocols (Ben-Porath, 1998, 2003; Gerardi, 2004) that provide implementation only for the case $k=1$. However, whereas the above papers focused on implementation as a sequential equilibrium, our implementation is as a Nash equilibrium.

We now present our result. We say that a correlated strategy profile q is *k-strong punishable* if there exists an uncorrelated strategy profile q' that is dominated by q for all players, even when a coalition with up to k players jointly deviate from q' (see Def. 2.4). Our result is the existence of a cheap-talk protocol that, for every $k<n/2$, implements any k-strong correlated equilibrium (with rational parameters), which is k-strong punishable, as a k-strong Nash equilibrium in the extended game with cheap-talk. Furthermore, we prove that in general such a protocol does not exist when $k \geqslant n/2$.

---

[2] Alternative definitions can be found in Milgrom and Roberts (1996), Ray (1996, 1998), Einy and Peleg (1995), and Bloch and Dutta (2009).

[3] An exception is a deviation of the grand coalition that may be easily coordinated (as there is no need to hide it from non-deviators). However, players are less concerned about such a deviation, because everyone earns from it.

[4] Those concepts somewhat resemble Eliaz's (1999) concept of k-Fault-Tolerant Nash Equilibrium.

The paper is organized as follows: Section 2 presents the model and formal definitions. Section 3 presents the main result. Section 4 shows that there are no similar protocols when $k \geq n/2$. Section 5 gives an example for the applicative use of our protocol. We conclude in Section 6.

2. Model and Definitions

A finite game in strategic form *G* is defined as $G = (N, (A^i)_{i \in N}, (u^i)_{i \in N})$, where $N = \{1,..,n\}$ is a non-empty finite set of players, and for each $i \in N$, A^i is player *i*'s non-empty finite set of pure actions, and u^i is player i's payoff function, a real-valued function on $A = \prod_{i \in N} A^i$. The multi-linear extension of u^i to $\Delta(A)$ is still denoted by u^i. A member of $\Delta(A)$ is called a (correlated) *strategy profile*. A coalition *S* is a non-empty member of 2^N. Given a coalition $S \subseteq N$, let $A^S = \prod_{i \in S} A^i$, and let $-S = \{i \in N \mid i \notin S\}$ denote the complementary coalition. A member of $\Delta(A^S)$ is called an *S-strategy profile*. Given $q \in \Delta(A)$ and $a^S \in A^S$, we define $q_{|S} \in \Delta(A^S)$ to be $q_{|S}(a^S) = \sum_{a^{-S} \in A^{-S}} q(a^S, a^{-S})$, and for simplicity we omit the subscript: $q(a^S) = q_{|S}(a^S)$. We say that $q \in \Delta(A)$ is an *uncorrelated strategy profile* if for every $a = (a^1,...,a^n) \in A$, $q(a) = q(a^1) \cdot ... \cdot q(a^n)$. Similarly, given $S \subseteq N$, we say that $q^S \in \Delta(A^S)$ is an *uncorrelated S-strategy profile* if for every $a^S \in A^S$, $q^S(a^S) = \prod_{i \in S} q^S(a^i)$. Let $IA \subset \Delta(A)$ be the set of uncorrelated strategy profiles, and let $IA^S \subset \Delta(A^S)$ be the set of uncorrelated *S*-strategy profiles. Given $q \in IA$, we write $q = (q^S, q^{-S})$ where: $q^S \in IA^S$, $q^{-S} \in IA^{-S}$.

Definition 2.1: An uncorrelated strategy profile $q \in IA$ is a *k-strong Nash equilibrium* if and only if for all coalitions $S \subseteq N$ satisfying $|S| \leq k$, and for every uncorrelated *S*-strategy profile $p^S \in IA^S$, there exists a player $i \in S$ such that $u^i(q) \geq u^i(p^S, q^{-S})$. Observe that a 1-strong Nash equilibrium is a Nash equilibrium (Nash, 1951), an *n*-strong Nash equilibrium is a strong Nash equilibrium (Bernheim et al., 1987), and any (*k*+1)-strong Nash equilibrium is also a *k*-strong Nash equilibrium.

Definition 2.2: Let $S \subseteq N$ be a coalition. An *S-deviating scheme* is a function $d^S : A^S \to \Delta(A^S)$. A strategy profile $p \in \Delta(A)$ is an *S-deviation* from the strategy profile $q \in \Delta(A)$, if there exists a deviating scheme d^S, such that for all $a \in A$, we have $p(a) = \sum_{b^S \in A^S} q(b^S, a^{-S}) \cdot d^S(a^S \mid b^S)$. Let $D(q,S) \subseteq \Delta(A)$ denote the *set of all S-deviations from q.*

Thus, a correlated strategy profile is an *S*-deviation from a given agreement *q* (a correlated strategy profile), if the members of *S*, using some plan to correlate their play, can induce the correlated strategy profile *p* when each member of the complementary coalition obeys the agreement.

Definition 2.3: A profile $q \in \Delta(A)$ is a *k*-strong correlated equilibrium if for every coalition $S \subseteq N$ satisfying $|S| \leq k$, and for each *S*-deviation $p \in D(q,S)$, there is a player $i \in S$, s.t. $u^i(q) \geq u^i(p)$.

A *k*-strong correlated equilibrium is a correlated strategy profile, from which no coalition, with up to *k* players, has a joint deviation, which makes every member of the coalition better off. Observe that a 1-strong correlated equilibrium is a correlated equilibrium, an *n*-strong correlated equilibrium is a strong correlated equilibrium (Moreno & Wooders, 1996), and any (*k*+1)-strong correlated equilibrium is also *k*-strong. Similar to the existing definitions of strong equilibria, we assume that deviating players are myopic: they do not take into account the possibility that there may be further deviations.[5]

Definition 2.4: Let $q \in \Delta(A)$ be a strategy profile. An uncorrelated strategy profile $\tilde{q} \in IA$ is a *k-strong punishing strategy profile* (for *q*) if for every coalition $S \subseteq N$ satisfying $|S| \le k$ and for every *S*-deviation $\tilde{p} \in D(\tilde{q}, S)$ (from $\tilde{q}$) there exists $i \in S$, such that $u^i(q) > u^i(\tilde{p})$. A strategy profile *q* is *k-strong punishable* if there exists a *k*-strong punishing strategy profile (for *q*).

In our cheap-talk protocol players construct together a correlation device that recommends each player what to play according to a *k*-punishable strategy profile *q*. The punishing strategy profile $\tilde{q}$ is used to prevent a coalition *S* from using an *S-lie* (defined in Sec. 3): deviate while communicating with the members of $-S$ in order to change their recommended actions.

Definition 2.5: Let $S \subseteq N$ a coalition and *q* a *k*-strong punishable strategy profile with *a k*-strong punishing profile $\tilde{q}$. The *S-punishment* is: $m^S_{q,\tilde{q}} = \min_{\tilde{p} \in D(\tilde{q},S)} \max_{i \in S} \left(u^i(q) - u^i(\tilde{p}) \right)$. The *minimal punishment* is: $m_{q,\tilde{q}} = \min_{S \subseteq N, |S| \le k} \left(m^S_{q,\tilde{q}} \right)$. The *maximal profit* is: $w_q = \max_{i \in N, a \in A} \left(u^i(a) - u^i(q) \right)$. The *minimal detecting probability* $0 \le \lambda_{q,\tilde{q}} < 1$ is the solution in the interval $[0,1)$ to the equation $(1 - \lambda_{q,\tilde{q}}) \cdot w_q = \lambda_{q,\tilde{q}} \cdot m_{q,\tilde{q}}$.

Given a coalition *S* with up to *k* players and an *S*-deviation $\tilde{p}$ from the punishing profile $\tilde{q}$, one of the deviators looses $\max_{i \in S} \left(u^i(q) - u^i(\tilde{p}) \right) > 0$ if the profile *q* is replaced with $\tilde{p}$. The *S*-punishment $m^S_{q,\tilde{q}} > 0$ is the minimal such loss for all possible *S*-deviations $\tilde{p} \in D(\tilde{q}, S)$, and the minimal punishment $m_{q,\tilde{q}} > 0$ is the minimal *S*-punishment for all coalitions with up to *k* players. The expression w_q is the maximal profit a player may earn from replacing the profile *q* with another profile.

The properties of our protocol guarantee that any use of an *S*-lie is detected by a non-deviating player with probability $\lambda > \lambda_{q,\tilde{q}}$. If an *S*-lie is detected, then all the non-deviators play the punishing strategy profile $\widetilde{q}$, and the member of *S* deviate from $\widetilde{q}$ to an *S*-deviation $\widetilde{p} \in D(\widetilde{q}, S)$. In such a case there exists a deviator (say player *i*) that looses at least $m_{q,\tilde{q}}$. If the *S*-lie is undetected then the profit of each deviator is at most w_q. Thus, Def. 2.5 implies that player *i* looses (in expectation) if the *S*-lie is used.

[5] Assuming otherwise leads to the concepts of a coalition-proof Nash equilibrium (Bernheim et al., 1987) and of a correlated coalition-proof equilibrium (see Moreno & Wooders, 1996, and the references mentioned in footnote 2).

Definition 2.6: Let $G = (N, (A^i)_{i\in N}, (u^i)_{i\in N})$ be a game, and let M be a finite alphabet that contains the null message ϕ. The *extended cheap-talk game* $\bar{G} = \bar{G}(M)$ is the following game with two phases: *talk phase* and *play phase.* The talk phase includes infinite number of stages.[6] At each stage each player i simultaneously sends a message to each non-empty coalition $i \notin S$. The messages are taken from the alphabet M. In the play phase, each player i simultaneously chooses an action in A^i.

Let $H_{t_0} = \prod_{t<t_0} \prod_{i\in N} \prod_{\phi\neq S\subseteq N\setminus\{i\}} M$ be the set of t_0-period histories, $H^i_{t_0}$ the set of t_0-period information sets of player i: the part of the history that includes messages sent by player i or received by a coalition that includes player i. Let H_∞ be the set of infinite histories, and H^i_∞ the set of infinite information sets of player i. A (behavioral) i-strategy in $\bar{G}$ is a pair of measurable[7] functions $c^i = (f^i, g^i)$ where:

- $f^i : \bigcup_{t\in\mathbf{N}} H^i_t \to \Delta(M^{2^{n-1}-1})$ - player i's function for choosing the messages he sends each non-empty coalition that that does not include him.
- $g^i : H^i_\infty \to \Delta(A^i)$ - player i's function for choosing his action in the play phase.

We use the term *protocol* to denote an uncorrelated strategy profile in $\bar{G}$ (an n-tuple $c = (c^1, ..., c^n)$). Given a protocol c and $h \in H$, we refer to $(g^1(h^1), ..., g^n(h^n))$ as the *protocol's recommendations*.

Definition 2.7: Let G be a game, and $\bar{G}$ its cheap-talk extension. A protocol c is *finite* if there exists a random variable t_* with a finite expected value (which we call the protocol's *length*), such that for all $i \in N$ and for all $t > t_*$, $f^i_t = \phi$. Observe that only the players who follow c are bounded by the protocol finiteness. A deviating coalition S can continue to send non-null messages at stages after t_*.

Definition 2.8: Let G be a game, and $\bar{G}$ its cheap-talk extension. The probability $q_c \in \Delta(A)$ is the unique probability according to which actions are chosen at the play phase, if everyone plays in $\bar{G}$ according to c. Let $q \in \Delta(A)$ be a strategy profile. We say that a *protocol c implements q* if $q = q_c$.

Definition 2.9: Let G be a game, $\bar{G}$ its cheap-talk extension, c a protocol, and S a coalition. A protocol c_S is an *S-protocol-deviation* (from c) if for every $i \notin S$, $c^i = c^i_S$.

3. Minority-Proof Cheap-Talk Protocol

Theorem 3: Let G be a game with n players, $k < n/2$, and $q \in \Delta(A)$ a k-strong correlated equilibrium, which is k-strong punishable, with rational parameters (i.e. $q(a)$ is rational $\forall a \in A$). Then there exists a finite alphabet M (which depends on q) such that in the extended cheap-talk game $\bar{G}(M)$

[6] We defined the cheap-talk to be infinite in the spirit of Aumann & Hart (2003) who discuss 2-player games, and show that any artificial restriction on the length of the talk, limits the set of equilibria in the extended game due to terminal effects propagating backwards.

[7] H_t, H^i_t have the discrete topology and H_∞, H^i_∞ have the usual product topology (smallest σ-field containing all finite cylinders).

there is a finite protocol *c* that implements *q* and is a *k*-strong Nash equilibrium.

Proof of Theorem 3: We first give a constructive description of the protocol, and then prove that it is a *k*-strong Nash equilibrium. For simplicity of presentation, we assume *n* to be odd and $n = 2 \cdot k + 1$. The talk phase is divided into correlation phases and monitoring phases.

Each *correlation* phase is based on the *k-private protocol* presented in Ben-Or et al. (1988), which deals with fault-tolerant distributed computation. Their setup includes *n* players, where $k<n/2$, each holding a secret input $x_i \in \mathbf{Z}_p$ ($\mathbf{Z}_p$ is the finite field of integers modulo *p*) who compute *n* polynomials $\left(f_i(x_1,...,x_n)\right)_{i\in N}$ - the outputs. Their protocol, if followed by all players, allows the players to obtain simultaneously their outputs at the end of the protocol. Specifically, each player *i* obtains the value of $f_i(x_1,...,x_n)$, while not acquiring any information about the values of the other outputs or inputs: the conditional distribution of $\left(x_j\right)_{j\in N, j\neq i}$ and $\left(f_j(x_1,...,x_n)\right)_{j\in N, j\neq i}$ given all the messages he received and sent (and his input x_i) is the same as the conditional distribution given only $f_i(x_1,...,x_n)$ and x_i.

The *k*-privacy property of the protocol means that if any coalition *S*, with up to *k* players, shares after the protocol ends all the messages each of them received and sent (and their inputs $\left(x_j\right)_{j\in S}$), then the resulting conditional distribution of $\left(f_j(x_1,...,x_n)\right)_{j\in -S}$ and $\left(x_j\right)_{j\in -S}$ is the same as the conditional distribution given only $\left(f_j(x_1,...,x_n)\right)_{j\in S}$ and $\left(x_j\right)_{j\in S}$. Moreover, if such sharing is done before the protocol ends, then the members of *S* do not acquire any information about any of the outputs $\left(f_i(x_1,...,x_n)\right)_{i\in N}$. An additional property of the protocol that will be useful later in the proof is the following: every coalition *S* with at-least *k*+1 players can share the messages received from each player $i \notin S$ at the first stage and reveal his input x_i.

Let $d \in \mathbf{N}$ be the common denominator of $\{q(a)\}_{a\in A}$, let $p \in \mathbf{N}$ be a prime number satisfying: $p > d$, $p > 1/\left(1-\lambda_{q,\tilde{q}}\right)$ and $\forall i \in N,\ p > |A_i|$, let the alphabet of $\bar{G}$ be $M = \mathbf{Z}_p \cup \phi$, and let $\left(f_i(x)\right)_{i\in N}$ be polynomials over $\mathbf{Z}_p$ which satisfy the following conditions: if x is chosen uniformly over $\{1,..,d\} \subseteq \mathbf{Z}_p$, then $\Pr\left(f_1(x),...,f_n(x) = \left(i_1,...,i_n\right)\right) = q(a_{i_1},...,a_{i_n})$, where $a_{i_j} \in A^j$ is the i_j-th action of player j; if $x \in \{d+1,...,p-1\}$, then $f_1(x) = ... = f_n(x) = 0$.

At the beginning of each correlation phase, each player randomly chooses a secret input $x_i \in \mathbf{Z}_p$ (according to the uniform distribution). The players communicate using the *k*-private protocol until each player *i* simultaneously obtains the value of $f_i(x) = f_i(x_1 + ... + x_n)$, which is interpreted to be the protocol's recommendation for player *i*: if $f_i(x) = l$ then his recommended action is the l-th action. If a

player receives an invalid message during the correlation phase (for example, he receives a null message instead of a number in $\mathbf{Z}_p$), then he sends null messages for the rest of the correlation phase. If some player i has not received a valid recommended action or if $f_i(x) = 0$, then he considers his recommended (mixed) action to be $\tilde{q}^i$, where $\tilde{q}$ is a k-strong uncorrelated punishing profile of q.

The following joint lottery and monitoring phase are based on Ben-Porath (1998). Each player i is being asked whether $f_i(x) = 0$. If all players give a positive answer, then a *monitoring* phase is executed, in which each player announces all the messages he sent and received in the last correlation phase and a new correlation phase is played. Otherwise, the players conduct a joint lottery:[8] with large enough probability $\lambda_{q,\tilde{q}} < \lambda < 1$, the monitoring phase is executed, and with probability 1-λ nothing is revealed, the talk phase ends (i.e. since that stage, everyone sends null messages), and each player plays his recommended action. If during the monitoring phase a deviation is revealed (i.e. a sender claims he sent some message while a receiver claims he received a different message, or some player i did not act according to the protocol, like sending a null message instead of a number in $\mathbf{Z}_p$.), then the players play the punishing profile $\tilde{q}$. Observe that if everyone follows the protocol, then it implements q.[9]

Proving the protocol is a *k*-strong Nash equilibrium:

Let S be a coalition with up to k players. We have to show that there is no profitable deviation for S, i.e. that for every S-protocol-deviation c_S there is a deviating player $i \in S$, such that $u^i(q_{c_S}) \le u^i(q_c)$. The possible S-protocol-deviations can be divided into a few kinds: choosing the inputs non-uniformly, sharing information, not following S-part of the action profile, and S-lies. We show that none of those kinds (nor a combination of them) is profitable for S.

Choosing inputs non-uniformly: In the beginning of the correlation phase, each player should uniformly choose a secret input $x_i \in \mathbf{Z}_p$. The fact that the action profile depends only on the sum of those x_i-s, guarantee that choosing x_i in any arbitrary way by the members of S, does not affect the distribution of the action profile.

Sharing information: In this deviation, S members follow the protocol c when communicating with members of $-S$, but deviate when communicating among themselves: They send messages that contain information about their secret inputs or about messages they received or sent in earlier parts of the protocol. Such sharing can be done in "silent" stages (when the players are supposed to send null

[8] The joint lottery is conducted by each player i simultaneously and publicly announces a random number $y_i \in \mathbf{Z}_p$ (chosen by the uniform distribution). The players play the monitoring phase if $y_1+\ldots+y_n \neq 0$. If some player i does not announce a number, it is assumed that $y_i=0$.

[9] A more detailed description of our protocol can be found in Heller (2008).

messages).[10] However, the *k*-privacy property of the protocol (described above) guarantees that such information sharing does not affect the conditional distribution of $\left(f_j(x)\right)_{j\in -S}$ and $\left(x_j\right)_{j\in -S}$.

Not following the protocol's recommendations: The players of *S* may plan a deviation in the playing phase: playing according to an *S*-deviating scheme $d^S : A^S \to \Delta(A)$ instead of following the protocol's recommendations. However, the fact that q is a *k*-strong correlated equilibrium and that $q = q_c$ guarantees that such a deviating scheme is not profitable for at least one player in *S*.

S-lies: We define an *S*-lie as an *S*-protocol-deviation in which the players of *S* deviate while communicating with the non-deviating players of $-S$, and as a result a non-deviating player $j \notin -S$ receives a different recommendation than $f_j(x)$ (or does not receive a valid recommendation at all). With probability λ a monitoring phase is executed after the correlation phase. As part of the monitoring, the members of $-S$ share the messages they received at the first stage of the *k*-private protocol. This allows each non-deviating player to evaluate the true value of all the inputs[11] (due to the last property of the *k*-private protocol described above), and to check whether any non-deviating player *j* received a different recommendation than $f_j(x)$. Thus any *S*-lie is detected with probability $\lambda > \lambda_{q,\tilde{q}}$, and Def. 2.5 guarantees that any such *S*-lie is not profitable to at least one player in *S*.[12] **QED**

4. Non-Existence of a Cheap-Talk (*n*/2)-Proof Protocol

In this section we show that our result is tight. Specifically, example 4 shows that for every *n,* there exists a game *G* with 2·*n* players and an *n*-strong correlated equilibrium $q \in \Delta(A)$ with rational parameters, which is *n*-strong punishable, such that no protocol in $\overline{G}$ that implements *q* is an *n*-strong Nash equilibrium.

Example 4: Let *G* be a game with 2·*n* players: $\{A^1,..., A^n, B^1, ..., B^n\}$. Each player in $A=\{A^1,...,A^n\}$ has two pure actions: $\{e^1,d^1\}$, and each player in $B=\{B^1,..., B^n\}$ has two pure actions: $\{e^2, d^2\}$. The payoff of the game is:

- If any two players in *A* played a different action (i.e. A^i played e^1 while A^j played d^1), or any two players in *B* played a different action, then all the players get 0.
- Otherwise (all players in each group play the same action), the payoff is as described in table 4.1.

[10] For example, the deviating players can use some of the infinite number of stages after the protocol ends for sharing information.
[11] We interpret the true value of the input of player *i* as the value implied by the messages sent to -*S* at the first stage. If no single value is implied, then it is considered as a deviation from the protocol (and the punishing strategy profile is used).
[12] Observe that the protocol is not a (*k*+1)-strong Nash equilibrium, as any coalition with *k*+1 players can share the messages received at the first stage of the *k*-private protocol and reveal the whole profile of recommendations.

Let q be the n-strong correlated equilibrium, which is n-strong punishable (with a punishing profile $(d^1,...,d^1,d^2,...,d^2)$), that is described in Table 4.2. Thus in q all players in A play the same action, as do all the players in B.

Table 4.1: *G's* payoff

		B: $(e^2,...,e^2)$	B: $(d^2,...,d^2)$
A	$(e^1,...,e^1)$	3 to everyone	1 to $\{A^1,...,A^N\}$ 4 to $\{B^1,...,B^N\}$
A	$(d^1,...,d^1)$	4 to $\{A^1,...,A^N\}$ 1 to $\{B^1,...,B^N\}$	0 to everyone

Table 4.2: *n*-strong correlated equilibrium *q*

		B: $(e^2,...,e^2)$	B: $(d^2,...,d^2)$
A	$(e^1,...,e^1)$	$1/3$	$1/3$
A	$(d^1,...,d^1)$	$1/3$	0

Assume to the contrary that there is a finite cheap-talk protocol c (with length t_*) such that $q=q_c$ and that c is an n-strong Nash equilibrium in $\overline{G}$. Let the history of messages (H_∞) be partitioned according to the messages transferred between members of A and members of B. Specifically, let $H_{\infty,A\leftrightarrow B}$ denote the part of history that includes messages sent by a player in A to a coalition that includes players in B, and messages sent by a player in B to a coalition that includes players in A. For each $h_{A\leftrightarrow B}\in H_{\infty,A\leftrightarrow B}$ let $H(h_{A\leftrightarrow B})$ denote the set of histories in which those transferred messages are equal to $h_{A\leftrightarrow B}$. Given $H'\subseteq H_{\infty,A\leftrightarrow B}$, let $H(H')=\bigcup_{h_{A\leftrightarrow B}\in H'}H(h_{A\leftrightarrow B})$. Let $q_{h_{A\leftrightarrow B}}$ denote the probability distribution on profiles of actions of the players (when everyone follows c) conditional on $H(h_{A\leftrightarrow B})$, and let $q^A_{h_{A\leftrightarrow B}}$ and $q^B_{h_{A\leftrightarrow B}}$ denote the marginals of $q_{h_{A\leftrightarrow B}}$ on profiles of actions of members in A and in B respectively. Given $h_{A\leftrightarrow B}$ the behavior of members of A is independent of the behavior of members of B. Thus, $q_{h_{A\leftrightarrow B}}$ is a product of $q^A_{h_{A\leftrightarrow B}}$ and $q^B_{h_{A\leftrightarrow B}}$. Let H^A_0 (H^B_0) be the set of $h_{A\leftrightarrow B}\in H_{\infty,A\leftrightarrow B}$ such that $q^A_{h_{A\leftrightarrow B}}$ ($q^B_{h_{A\leftrightarrow B}}$) assigns a positive probability to profiles of actions where members of A (B) play different actions. Clearly $\Pr_c(H(H^A_0))=\Pr_c(H(H^B_0))=0$. Thus, we can assume that the players in each group play the same action: $q^A_{h_{A\leftrightarrow B}}\in\Delta\left(\left\{\vec{d}^1,\vec{e}^1\right\}\right)=\Delta\left(\left\{\left(d^1,...,d^1\right),\left(e^1,...,e^1\right)\right\}\right)$ and $q^B_{h_{A\leftrightarrow B}}\in\Delta\left(\left\{\vec{d}^2,\vec{e}^2\right\}\right)$. We now show that with probability 1 both $q^A_{h_{A\leftrightarrow B}}$ and $q^B_{h_{A\leftrightarrow B}}$ are pure:

- Let H_I be the set of $h_{A\leftrightarrow B}\in H_{\infty,A\leftrightarrow B}$ where both $q^A_{h_{A\leftrightarrow B}}$ and $q^B_{h_{A\leftrightarrow B}}$ are mixed. Clearly $\Pr_c(H(H_I))=0$ (because otherwise $q_c\left(\vec{d}^1,\vec{d}^2\right)>0$).
- Let H^A_{II} (H^B_{II}) be the set of $h_{A\leftrightarrow B}\in H_{\infty,A\leftrightarrow B}$ where $q^A_{h_{A\leftrightarrow B}}$ ($q^B_{h_{A\leftrightarrow B}}$) is mixed while $q^B_{h_{A\leftrightarrow B}}$ ($q^A_{h_{A\leftrightarrow B}}$) is pure. In every $h\in H(H^A_{II})$ ($h\in H(H^B_{II})$), the members of A (B) has a profitable deviation: if B's (A's) action is $\vec{d}^2$ ($\vec{d}^1$), then the members of A (B) deviate and play $\vec{e}^1$ ($\vec{e}^2$), and if B's (A's) action is $\vec{e}^2$ ($\vec{e}^1$), then the members of A (B) deviate and play $\vec{d}^1$ ($\vec{d}^2$). Thus, $\Pr_c(H(H_{II}))=0$.

Let H_{III} be the set of $h_{A\leftrightarrow B}\in H_{\infty,A\leftrightarrow B}$ where $q^A_{h_{A\leftrightarrow B}}=\vec{e}^1$ and $q^B_{h_{A\leftrightarrow B}}=\vec{e}^2$. Observe that $\Pr_c(H(H_{III}))$

=1/3 (because $q_c(\vec{e}^1,\vec{e}^2)=1/3$). We finish the proof by observing that both groups have a profitable deviation: playing $\vec{d}^i$ instead of $\vec{e}^i$ in every history in $H(H_{III})$, contradicting our assumption.[13] **QED**

5. An Example for Applicability – a 5-Player "Chicken" Game

In this section, we study a 5-player "chicken" game, in which the use of our "minority-proof" protocol can give a substantial gain to all players. Let *G* be the following game:

- Each of the 5 players has two pure actions: *s* ("swerve") and *d* ("drive straight")
- The payoff function is:
  - If all players play *s*, then everyone gets 4.
  - If up to 2 players play *d*, then those who played *d* get 5 while the others get 2.
  - If more than 2 players play *d*, then everyone gets 0.

The presence of a fair mediator allows the players to achieve the following correlated strategy profile *q*:

- With probability 3/8: all the players play *s*.
- For each of the 10 couples (*i,j*) s.t. $i \neq j$, with probability 1/16 (*i,j*) play *d*, while the others play *s*.

One can verify that:

- The profile *q* is a 2-strong correlated equilibrium with a symmetric payoff of 3.5 which is the best 2-strong correlated equilibrium symmetric payoff.
- The profile *q* is 2-strong punishable (with the punishing strategy $\tilde{q}=(d,d,d,d,d)$).
- The payoff of *q* is strictly better than the best symmetric payoff in the convex hull of Nash equilibria (3.2, Achieved by choosing each of the ten couples (*i,j*) with probability 10%, and playing the Nash equilibrium in which (*i,j*) play *d* and the others play *s*.)

We now compare our protocol with the existing protocols in the literature (Forges, 1990; Barany, 1992; Ben-Porath, 1998, 2003; Gerardi, 2004), when the players use cheap-talk. These protocols implement *q*, but only as a (1-strong) Nash equilibrium. This implementation is "weak" in the sense that it is possible for two players to jointly deviate and guarantee a payoff of 5 for themselves (and 2 to the other players). Contrary to that, the use of our protocol gives a "stronger" implementation as a 2-strong Nash equilibrium. An analog example can be devised for any odd number of players.

6. Concluding Remarks:

1) ***n*/2-privacy and *n*/3-resiliency**: Ben-Or et al. (1988) present a distributed computation protocol for *n*

[13] The members of *A* (*B*) use stages after the original protocol ends (t_*) to share the information about $h_{A\leftrightarrow B}$. If in a different pre-play communication framework, players can limit the private communication channels of sub-coalitions (for example, at some point the grand coalition can decide that the talk phase ends, and the play phase immediately begins), then our proof does not hold.

players (as described in Sect. 3) with the following properties:

- $n/2$-privacy - If everyone follows the protocol, then no coalition with less than $n/2$ players can get any additional information about the outputs of the other players.
- $n/3$-resiliency - No coalition with less than $n/3$ players can either disrupt the computation or get additional information about the outputs of the other players.

The latter property directly implies that it is possible to use the protocol of Ben-Or et al. to implement any $n/3$-strong correlated equilibrium as an $n/3$-strong Nash equilibrium in an extended game with cheap-talk. The main contribution of this paper is to show that it is possible to implement an $n/2$-strong correlated equilibrium that is $n/2$-strong punishable, using the procedure of repeated random monitoring that was introduced in Ben-Porath (1998).

2) **Simultaneousness and private channels:** The protocol presented in this paper relies on simultaneous communication, and assumes that players can send and receive messages exactly at the same time. In some real-world environments this assumption does not hold. In Heller (2008) we relax this assumption (but strengthen the k-strong punishability requirement), and present a *polite* protocol that does not rely on simultaneous communication. Moreover, the protocol uses only 2-player private communication channels, and does not rely on public communication channels.

3) **Two possible extensions of our protocol are:**

- Implementing a k-coalition-proof correlated equilibrium as a k-coalition-proof Nash equilibrium (Bernheim et al., 1987) in the extended cheap-talk game.
- Implementing a k-strong (or k-coalition-proof) correlated equilibrium of a Bayesian game.

4) Cryptographic protocols: In situations where players are computationally restricted and one assumes existence of "one-way" functions, it is possible to construct a protocol that implements any k-strong correlated equilibrium as a k-strong Nash equilibrium, without the restriction $k < n/2$ (see: Gossner, 1998; Urbano and Vila, 2002; Lepinski et al., 2004; and Abraham et al., 2006).

References:

Abraham I., Dolev D., Gonen R., Halpern J., 2006. Distributed computing meets game theory: robust mechanisms for rational secret sharing and multiparty computation. Proc. 25 ACM , 53-62.

Aumann, R., 1959. Acceptable points in general cooperative n-person games, in Kuhn, H.W., Luce, R.D. (Eds.), Contributions to the theory of games IV. Princeton University Press, N.J, pp. 287-324.

Aumann, R., 1974. Subjectivity and correlation in randomized strategies. J. Math. Econ. 1, 67-96.

Aumman, R., Hart S., 2003. Long cheap talk. Econometrica 71 (6), 1619-1660.

Barany, I., 1992. Fair distribution protocols or how the players replace fortune. Mathematics of Operations Research. 17, 329-340.

Ben-Or, M., Goldwasser S., Wigderson A., 1988. Completeness theorems for non-cryptographic fault-tolerant distributed computation (extended abstract). Proc. 20 STOC, ACM, 1-10.

Ben-Porath, E., 1998. Communication without mediation: expending the set of equilibrium outcomes by "cheap" pre-play procedures. J. Econ. Theory 80, 108-122.

Ben-Porath, E., 2003. Cheap talk in games with incomplete information. J. Econ. Theory 108, 45-71.

Bernheim, B., Peleg, B., Whinston, M., 1987. Coalition-proof Nash equilibria. J. Econ. Theory 42, 1-12.

Bloch F., Dutta B., 2009. Correlated equilibria, incomplete information and coalitional deviations. ames Econ. Behav. 66, 721-728.

Crawford, V., Sobel, J., 1982. Strategic information transmission. Econometrica 50, 579-594.

Einy, E., Peleg B., 1995. Coalition proof communication equilibria. In: Social Choice, Welfare & Ethics (W. Barnett, H. Moulin, M. Salles and N. Schofield, eds.), Cambridge, New-York and Melbourne.

Eliaz K., 1999. Fault tolerant implementation. Review of Economic Studies, Vol. 69(3), 589-610.

Forges, F., 1990. Universal mechanisms. Econometrica 58, 1341-1364.

Gerardi, D., 2004. Unmediated communication in games with complete and incomplete information. J. Econ. Theory 114, 104-131.

Gossner, A., 1998. Secure protocols or how communication generates correlation. J. Econ. Theory 83, 69-89.

Heller Y., 2008, A minority-proof cheap-talk protocol - an extended version, mimeo, http://www.tau.ac.il/~helleryu/minority-extended.pdf

Lepinski, M., Micali, S., Peikert C., Shelat A., 2004. Completely fair SFE and coalition-safe cheap talk. Proc. 23 ACM, 1-10.

Milgrom, P., Roberts, J., 1996. Coalition-proofness and correlation with arbitrary communication possibilities, Games Econ. Behav. 17, 113-128.

Moreno, D., Wooders, J., 1996. Coalition-proof equilibrium, Games Econ. Behav. 17, 80-113.

Molotov-Ribbentrop Pact, 1939. Modern history sourcebook edited by Paul Halsall (1997), http://www.fordham.edu/halsall/mod/1939pact.html (visited in November 2008).

Nash, J.F., 1951. Non-cooperative games. Ann. Math. 54, 286-295.

Ray I., 1996. Coalition-proof correlated equilibrium: a definition. Games Econ. Behav. 17, 56-79.

Ray, I., 1998. Correlated equilibrium as a stable standard of behavior. Rev. Econ. Design, 3, 257-269.

Urbano, A., Vila J.E, 2002. Computational complexity and communication: coordination in two-player games. Econometrica 70 (5), 1893-1927.