Economics 8117-8 **Noncooperative Game Theory**

November 25, 1997 **Lecture 12**

Professor Andrew McLennan

Coalition Proof Equilibrium and Correlated Equilibrium

I. Underlying Motivation of Coalition Proof Equilibrium.

A. The players can communicate, publicly and privately, before the play of the game.

B. There is no mechanism for enforcing agreements, so binding commitments are impossible.

II. The first concept proposed for this situation was Aumann's notion of strong equilibrium.

A. A *strong equilibrium* is a vector of mixed strategies with the property that no coalition can deviate in a way that gives each deviator a higher expected utility.

1. It is easy to find examples for which there are no strong equilibria.
2. A strong equilibrium is clearly highly stable against "conspiracies."

B. This notion can be criticized on the following counts.

1. There may not be any strong equilibria in two player games that are dominance solvable, for instance the Cournot game.
2. The equilibrium condition treats the equilibrium position and the deviations asymmetrically. The equilibrium must be stable against all coalitions, but the deviations do not need to be stable against anything.

III. Definitions.

A. Let $\mathcal{J} = \{J \subset I | \emptyset \neq J \neq I\}$ be the set of nonempty proper subcoalitions.

B. For $J \in \mathcal{J}$ and $\sigma_J = (\sigma_i)_{i \in J} \in \Pi_{i \in J}\, \Delta(S_i)$, let

$$N_{-J}(\sigma_J) = ((S_i)_{i \in I-J},\ (u_i(\cdot, \sigma_J))_{i \in I-J})$$

be the game with agents $I - J$, the original strategies for these agents, and the obvious utility functions.

Definitions:

(a) If $^{\#}I = 1$, $\sigma^* \in \Sigma$ is a *coalition–proof Nash equilibrium* if σ^* is an optimal strategy for the agent.

(b) If $^{\#}I > 1$, $\sigma^* \in \Sigma$ is *self–enforcing* if, for all $J \in \mathcal{J}$, σ_J^* is a coalition–proof Nash equilibrium of $N_{-(I-J)}(\sigma_{I-J}^*)$.

(c) If $^{\#}I > 1$, $\sigma^* \in \Sigma$ is a *coalition–proof Nash equilibrium* if it is self–enforcing and not strongly Pareto dominated in the class of self–enforcing strategy vectors.

IV. Nonexistence

A. Consider the game

	F			B	
	L	R		L	R
U	b	b	U	a	b
D	b	a	D	b	b

B. Here the three agents choose simultaneously, with 1 choosing between U and D, 2 choosing between L and R, and 3 choosing between F and B.

C. There are two outcomes. Assume that $a \succ_1 b$, $a \succ_2 b$, and $a \prec_3 b$.

D. It is easily verified that (U, L, F), (D, R, B), and $(\frac{1}{2}U + \frac{1}{2}D, \frac{1}{2}L + \frac{1}{2}R, \frac{1}{2}F + \frac{1}{2}B)$ are the only Nash equilibria.

E. It is also easy to see that in all three of these equilibria agents 1 and 2 can profitably conspire against agent 3.

V. Introduction to Correlated Equilibrium.

A The notion of correlated equilibrium is motivated by the following scenario.

1. A group of n agents must play a normal form game $(S_1, \ldots, S_n; u_1, \ldots, u_n)$.
2. Prior to the play of the game they can negotiate, publicly, about how the game should be played.
3. They cannot make agreements that are binding in the sense of providing some method (outside the given game) for punishing agents who violate the agreement.

4. They do have access to an impartial mediator who can make private recommendations according to any agreed (random) pattern.

B. No attempt is made to predict which outcome will actually be chosen.

1. Such predictions are the province of bargaining theory and/or cooperative game theory.
2. Rather, the idea is to determine the set of agreements that are plausible.
3. In modern terminology the restrictions on possible agreements are called incentive compatibility constraints.

VI. Formal Development

A. **Definition:** $\mu \in \Delta(S)$ is a *correlated equilibrium* if, for all $i = 1, \ldots, n$ and all $s_i, t_i \in S_i$,

$$(*) \qquad \sum_{s_{-i} \in S_{-i}} \mu(s_i, s_{-i}) u_i(s_i, s_{-i}) \geq \sum_{s_{-i} \in S_{-i}} \mu(s_i, s_{-i}) u_i(t_i, s_{-i})$$

1. The idea is that, with probability $\mu(s)$, the mediator recommends the various components of s to the agents.
2. Condition $(*)$ says that, when s_i is recommended to agent i, on average agent i does no better by deviating to t_i.

Theorem 1: The set of correlated is a convex polytope in $\Delta(S)$.

Proof: By definition a convex polytope is the intersection of finitely many half spaces. The relevant half spaces are given by the conditions $\mu(s) \geq 0, s \in S, \Sigma_{s \in S}\, \mu(s) \geq 1, \Sigma_{s \in S}\, \mu(s) \leq 1,$ and the inequalities $(*)$ for the various i and s_i, $t_i \in S_i$. ∎

Theorem 2: If $\mu = \sigma_1 \times \ldots \times \sigma_n$, where $\sigma = (\sigma_1, \ldots, \sigma_n) \in \Sigma$, then μ is a correlated equilibrium if and only if σ is a Nash equilibrium. Consequently the set of correlated equilibria is nonempty.

Proof: When $\mu = \sigma_1 \times \ldots \times \sigma_n$ condition $(*)$ reduces to

$$0 \le \sum_{s_{-i} \in S_{-i}} \mu(s_i, s_{-i})[u_i(s_i, s_{-i}) - u_i(t_i, s_{-i})]$$
$$= \sigma_i(\sigma_i)[u_i(s_i, \sigma_{-i}) - u_i(t_i, \sigma_{-i})]. \quad \blacksquare$$

Remark: In words, Theorem 2 says that the Nash equilibria are precisely the uncorrelated correlated equilibria.

Theorem 3: There may be correlated equilibria which Pareto dominate all Nash equilibria. Consequently there may be correlated equilibria that are not convex combinations of Nash equilibria.

Proof: Consider the following game, which expands on an example due to Aumann:

1\2	L	M	R
U	$(6, 6)$	$(1, 7)$	$(-10, -10)$
C	$(7, 1)$	$(0, 0)$	$(-10, 2)$
D	$(-10, -10)$	$(2, -10)$	$(-10, -10)$

The only Nash equilibrium not using D or R is $(\frac{1}{2}U + \frac{1}{2}C, \frac{1}{2}L + \frac{1}{2}M)$ with expected payoffs $(3\frac{1}{2}, 3\frac{1}{2})$. There is the quite unpleasant equilibrium (D, R) with expected payoffs $(-10, -10)$, and the other pure equilibria are (D, M) and (C, R). In a mixed equilibrium in which agent 1 assigns positive probability to D it must be the case that M is much more probable than L, which implies that no probability is assigned to C, which implies that no probability is assigned to L (unless agent 1 is always playing D), which implies that agent 1 assigns all probability to D, so that agent 2's expected payoff is -10 and agent 1's expected payoff is no greater than 2. A similar argument pertains if agent 2 is assigning positive probability to R.

The expected payoffs of the correlated equilibrium

$$\begin{array}{cc} 1\backslash 2 & \begin{array}{ccc} L & M & R \end{array} \\ \begin{array}{c} U \\ C \\ D \end{array} & \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & 0 \\ \frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \end{array}$$

are $(4\frac{2}{3}, 4\frac{2}{3})$. ∎

VII. Remarks.

A. The mediator need not be a person. It could be a programmable machine. Another point of view, discussed at length by Aumann ("Correlated Equilibrium as an Expression of Bayesian Rationality," *Econometrica*, **55**, (1987), 1-18) is that there is a probability space $(\Omega, \mathcal{A}, P)$ of payoff–irrelevant states, and each agent's information is described by a sub σ–algebra $\mathcal{A}_i$ of $\mathcal{A}$, so that each agent's strategy can be any $\mathcal{A}_i$–measurable function $\rho_i : \Omega \to \Delta(S_i)$. In this context correlated equilibrium can be derived from simultaneous satisfaction by all agents of classical axioms of Bayesian statistical decision theory.

B. In practice correlated equilibria do not seem to be very popular as outcomes of negotiations. (In contrast randomizations over Nash equilibria – "Let's flip a coin" – are commonplace.) Various factors can be cited – possibilities of side payments, expenses of human or computerized mediation – but there may also be difficulties with the concept that are as yet unmodelled.

C. To my knowledge it has not been done, but it is conceptually sensible to define hybrid concepts combining correlated equilibrium and coalition proof equilibrium. A variety of concepts may be possible, with the concepts differentiated according to the types of correlation and communication devices available to the coalitions.