**Economics 8117–8** **Noncooperative Game Theory**

**April 23, l996** **Lecture 26**

**Professor Andrew McLennan**

## Signalling Games

**I.** Introduction.

A. A signalling game consists of the following three phases.

1. Player $A$'s *type* is selected by nature.
2. Informed of his type, $A$ chooses a *message* $m \in M$.
3. Observing the message but not the type, player $B$ chooses a *response* $r \in R$.
4. The utilities for $A$ and $B$ are $u(t, m, r)$ and $v(t, m, r)$ respectively.

B. Signalling games have been analyzed in several contexts.

1. Principal–agent models.
2. Warranties.
3. Bargaining with incomplete information.
4. Entry deterence.
5. The chain–store game.

C. They provide a simple setting in which solution concepts can be tested.

1. Further refinements beyond sequential equilibrium have bite in this context.
2. The consequences of some concepts can be worked out completely.

D. In particular, Cho and Kreps are interested in determining the extent to which the consequences of stability are intuitive.

1. They are especially concerned with generalizations and variants of the following idea: if there are only two types and one loses (relative to equilibrium payoffs) if he chooses a particular message, no matter what the response, while the other gains if, after that message is chosen, $B$'s beliefs assign high probability to that type, then the equilibrium is not reasonable since the type who would

gain by sending the message should be able to do better than his supposed equilibrium payoff.

2. Several such conditions exist in the literature.
   a. Banks and Sobel.
   b. Grossman and Perry.
3. It should be emphasized that all these analyses are restricted to generic signalling games. In particular no one has analyzed the very interesting case in which the message does not affect payoffs and is thus “pure communication”.

**II.** A Review of Important Facts.

A. **Fact 1:** The set of Nash (and also sequential) equilibria of any finite player finite pure strategy noncooperative game consists of a finite number of connected sets.

B. **Fact 2:** For generic extensive games, the set of Nash equilibrium outcomes (probability distributions on terminal nodes) is finite.

C. **Fact 3:** For generic extensive games, a single equilibrium outcome is associated with each connected set of Nash (sequential) equilibria.

D. A set $M$ of Nash equilibria is *stable* if for every $\varepsilon > 0$ there is $\delta > 0$ such that if, for each $i$, $\sigma_i^*$ is given and $0 < \delta_i < \delta$, then the $\varepsilon$–ball around $M$ contains an equilibrium of the game in which the payoff associated with the strategy vector $s = (s_i)$ is replaced by the payoff associated with the mixed strategy vector $(\delta_i \sigma_i^* + (1 - \delta_i)s_i)_{i \in I}$.

1. Cho and Kreps do not require that a stable set be minimal with respect to this property.
2. **Fact 4:** For any finite game one of the components of the set of equilibria is a stable set.
3. **Fact 5:** Suppose we are given a stable set of equilibria for a game and a particular pure strategy for a given player $i$ that is never a weak best response

for the player relative to the set. Consider the game with this pure strategy removed (from the normal form) entirely, and consider the subset of the original stable set that consists of all strategy profiles from the set in which the player puts zero weight on this strategy. This subset is stable in the game that results after the particular strategy for the player is "pruned." The same is true if the strategy for the player that is pruned is weakly dominated by some other strategy for the player.

E. **Lemma 1:** Fix a signalling game. Suppose that an equilibrium outcome is stable, in the sense that the set of Nash equilibria that give rise to the outcome is a stable set. Then the set of all sequential equilibria that give rise to the outcome is also a stable set. Also, every stable set of equilibria contains at least one sequential equilibrium.

**Proof:** The subset of the stable set consisting of those equilibria that are perfect is also a stable set. (This is an immediate consequence of the definition of stability.) Since the normal form and the agent normal form coincide for signalling games, normal form perfect equilibria are agent normal form perfect and thus sequential. ■

**III.** Notation.

A. The Given Elements.

1. A *prior* $\pi \in \Delta(T)$.
2. For each $t \in T$, the set $M(t) \subset M$ of *available* or *feasible* messages.
   a. For $m \in M$ let $T(m) = \{t \in T | m \in M(t)\}$.
3. For each $m \in M$, the set $R(m) \subset R$ of *available* or *feasible* responses.

B. Strategies and Beliefs.

1. A strategy for $A$ is $\rho \in \Delta(M)^T$ where each $\rho(t) \in \Delta(M(t))$. We write $\rho(m;\, t)$ for the probability of $m$ given $t$.
2. Without great difficulty it can be shown that the only consequence of consis-

tency in this framework is Bayes rule:

$$\mu(t|m) = \frac{\pi(t)\rho(m;\, t)}{\Sigma_{t'}\, \pi(t')\rho(m;\, t')} \quad \text{if} \quad \Sigma_{t'}\, \pi(t')\rho(m;\, t') > 0.$$

3. A strategy for $B$ is $\varphi \in \Delta(R)^M$ where each $\varphi(m) \in \Delta(R(m))$. We write $\varphi(r;\, m)$ for the probability of $m$ given $t$.

C. Best responses.

1. For $m \in M$ and $\mu \in \Delta(T(m))$,

$$BR(\mu,\, m) = \text{argmax}_{r \in R(m)}\, \Sigma_{t \in T(m)}\, v(t,\, m,\, r)\mu(t).$$

2. For $m \in M$ and $I \subset T(m)$, $BR(I\,, m) = \bigcup_{\mu \in \Delta(I)} BR(\mu,\, m)$.

3. Let $MBR(\mu,\, m) = \Delta(BR(\mu;\, m))$, and let $MBR(I,\, m) = \Delta(BR(I;\, m))$.

**IV.** Methodology and Results.

A. Several concepts are defined in the following way.

1. In each there is, for each $m$, a set $T^s(m)$ (that may depend on the set of equilibria, usually a component, under consideration) of types that could "reasonably" choose $m$.

2. If there are no sequential equilibria in the given set with beliefs concentrated on the sets $T^s(m)$, then that set of equilibria is discarded as unreasonable.

B. Weak Dominance.

1. Here $T^s(m) = \{t|m$ is not a weakly dominated choice for $t\}$.

2. In view of Fact 5, for every stable set of sequential equilibria the subset consisting of those equilibria with beliefs concentrated on the sets $T^s(m)$ is stable for the game obtained by replacing $T(m)$ with $T^s(m)$.

C. The Intuitive Criterion.

1. A message is *out of equilibrium,* relative to a given equilibrium $(\rho,\, \varphi)$, if $\rho(m;\, t) = 0$ for all $t \in T(m)$. For each out of equilibrium message let

$$S(m) = \left\{ t|u^*(t) > \max_{r \in BR(T(m),\, m)} u(t,\, m,\, r) \right\}.$$

a. Here $u^*(t)$ is $t$'s expected payoff in the equilibrium $(\rho, \varphi)$.

2. The equilibrium fails the *intuitive criterion* (and satisfies it otherwise) if there is an out of equilibrium message $m$ and $t' \in T(m)$ with

$$u^*(t') < \min_{r \in BR(T(m)-S(m),\, m)} u(t', m, r).$$

a. Equilibria that fail the intuitive criterion are "intuitively" unstable because choosing such a message is thought to implicitly transmit the following "speech": "Look, there is no way to explain a choice of $m$ by someone in $S(m)$, so you ought to regard that set as having posterior probability 0 if you can think of some better explanation of why $m$ might be chosen. And of course there is such an explanation, since $t'$ would benefit from choosing $m$ so long as your beliefs don't assign probability to $S(m)$."

3. The intuitive criterion depends only on the outcome associated with $(\rho, \varphi)$. (Actually it depends only on the expected payoffs.)

4. The outcome associated with a sequential equilibrium $(\rho, \mu, \varphi)$ with $\mu(S(m)|m) = 0$ for all $m$ necessarily satisfies the intuitive criterion.

5. By choosing trembles that assign very small involuntary probabilities to the events "$t$ chooses $m$ where $t \in S(m)$" we can insure that the associated trembling hand equilibria satisfy the intuitive criterion (in the appropriate approximate sense). Thus every stable set of outcomes includes outcomes satisfying the intuitive criterion.

D. Divinity.

1. For a fixed equilibrium outcome, an out of equilibrium message $m$, and each $t$, let

$$D_t = \{\varphi \in MBR(T(m), m) | u^*(t) < \Sigma_r\, u(t, m, r)\varphi(r)\}, \quad \text{and let}$$

$$D_t^0 = \{\varphi \in MBR(T(m), m) | u^*(t) = \Sigma_r\, u(t, m, r)\varphi(r)\}.$$

2. **Criterion** $D1$**:** If, for some type $t$, there exists $t'$ with $D_t \cap D_t^0 \subset D_{t'}$; then $(t, m)$ may be pruned from the game.

3. **Criterion** $D2$**:** If, for some type $t$, $D_t \cap D_t^0 \subset \bigcup_{t' \neq t} D_{t'}$, then $(t, m)$ may be pruned from the game.

4. **Proposition:** If either of these criteria allow $(t, m)$ to be pruned, then $m$ is never a weak best response for $t$ at the given outcome.

   **Proof:** If $m$ was a weak best response for $t$ then it would be a strict improvement over the supposed equilibrium action for some $t'$ under either $D1$ or $D2$.

5. Divinity is built up out of iterative applications of these kinds of eliminations.

E. Never a Weak Best Response.

1. If an action is not a weak best response at any (generally sequential) equilibrium in a component of the set of equilibria, then there are sequential equilibria with beliefs that, "to the extent possible," do not assign probability to the action.

2. If there are several such actions in an extensive form then one can create various criteria on beliefs or deal with equilibria of the truncated game.

3. This is similar to justifiability except that the criterion of "never a weak best response" is here applied to the equilibria in a single component of the set of equilibria, whereas justifiability asks whether the action is ever a weak best response in the set of all sequential equilibria.