Economics 8117-8 **Noncooperative Game Theory**

April 21, 1998 **Lecture 18**

Professor Andrew McLennan

Justifiability

I. Introduction

A. Kreps and Wilson point out that, since beliefs are explicit in sequential equilibrium, one may be able to define interesting refinements of sequential equilibrium by imposing "reasonable" or "intuitive" restriction on beliefs.

B. I define such a refinement in my 1985 paper "Justifiable Beliefs in Sequential Equilibrium."

1. The basic idea is to require agents' beliefs to conform to the (perhaps dubious) notion that one deviation from the equilibrium path is more likely than another deviation if the first can be explained in terms of some confusion concerning which sequential equilibrium is "in effect,", whereas the second deviation cannot be explained in these terms.

2. Many different versions of this notion are possible. All involve some ad hocery.

3. In the best known examples justifiability performs reasonably well.

II. Definitions and Results

A. Introduction.

1. The idea is to impose some restrictions on beliefs.

   a. These restrictions will have some intuitive plausibility.

   b. Sequential equilibria with such beliefs always exist.

2. Review of the Kohlberg example

   a. (R, r) is proper.

   b. (R, r) is not stable, but that fact is difficult to relate to our intuitive reasons for regarding this equilibrium with suspicion.

3. Outline of remainder.

a. We develop the notion of first order justifiability rigorously, proving existence.

b. We then describe how to extend the analysis to higher order justifiability.

c. The intuitive flaws of this concept are discussed.

d. We relate the concept to stability.

B. First order justifiability

1. Recall that if $a \in A(h)$, we define $\pi|a$ to be π with π_h replaced by δ_a, the probability measure on $A(h)$ that assigns probability 1 to a.

Definition: An action $a \in A(h)$ is *first order useless* if for all sequential equilibria (μ, π),

$$\mathbf{E}^{\mu, \pi|a}(u_{\iota(h)}|h) < \mathbf{E}^{\mu, \pi}(u_{\iota(h)}|h).$$

A belief μ is *first order justifiable* if for all h and all $x, x' \in h$, $\mu(x) = 0$ whenever the number of useless actions in $\{\alpha(p_0(x)), \ldots, \alpha(p_{\ell(x)-1}(x))\}$ exceeds the number of useless actions in $\{\alpha(p_0(x')), \ldots, \alpha(p_{\ell(x')-1}(x'))\}$ A sequential equilibrium (μ, π) is *first order justifiable* if μ is justifiable.

Proposition: There exists a first order justifiable equilibrium.

Proof: Let $L = \max_{t \in T} \ell(t)$ be the *height* of the tree. Let $\{\delta_r\}_{r=1,2,\ldots}$ be a sequence in $(0, 1)$ converging to 0, and for each r define $\varepsilon_r : A \to (0, 1)$ by letting $\varepsilon_r(a) = \delta_r^L$ if a is first order useless and letting $\varepsilon_r(a) = \delta_r$ otherwise. For each r let π_r be an ε_r–perfect equilibrium of the agent normal form, and let μ_r be the belief generated by π_r. Taking a subsequence if necessary, assume $(\mu_r, \pi_r) \to (\mu^*, \pi^*)$, which is, of course, a sequential equilibrium.

Fix h and $x, x' \in H$, and assume that

$$m(x) = {}^{\#}\{\ell|\alpha(p_\ell(x)) \text{ is first order useless}\} > m(x').$$

Then for large r,

$$\mathbf{P}^{\pi}(x) = \rho(p_{\ell(x)}(x)) \cdot \Pi_{\ell=0,\ldots,\ell(x)-1}\, \pi(\alpha(p_\ell(x))) \leq (\delta_r^L)^{m(x)} \quad \text{and}$$

$$\mathbf{P}^{\pi}(x') \geq \rho(p_{\ell(x')}(x'))(\delta_r^L)^{m(x')}\, \delta_r^{L-1}.$$

Therefore

$$\mu^*(x) = \lim \mu_r(x) \leq \lim \frac{\mu_r(x)}{\mu_r(x')} \leq \delta_r^{L(m(x)-m(x'))-(L-1)} \leq \delta_r \to 0. \quad \blacksquare$$

C. Second order justifiability

1. Consider the example of Figure 1.

2. **Definition:** An action $a \in A(h)$ is *second order useless* if it is not first order useless but

$$\mathbf{E}^{\mu,\,\pi|a}(u_{\iota(h)}|h) < \mathbf{E}^{\mu,\,\pi}(u_{\iota(h)}|h)$$

for all sequential equilibria with first order justifiable beliefs. Beliefs are *second order justifiable* if, in each information set, positive probability is assigned only to those nodes requiring the fewest first order useless actions to be reached, and, in that subset, only those nodes requiring the fewest second order useless actions. A *second order justifiable equilibrium* is a sequential equilibrium with second order justifiable beliefs.

Proposition: There exists a second order justifiable equilibrium.

Proof: Choosing a sequence $\{\delta_r\}$ as above, define ε_r by $\varepsilon_r(a) = (\delta_r)^{L^2}$ if a is first order useless, δ_r^L if a is second order useless, and δ_r otherwise. Proceed as above. ∎

D. Critique

1. Counting procedures are clearly ad hoc
   a. There are weaker forms of the justifiability conditions that are less vulnerable to this criticism.
      i. One could merely require that all belief probability be assigned to nodes that require no useless actions to be reached.
      ii. One can require that no probability is assigned to a node if there is another node in the same information set that is reached through a set of useless actions that is a proper subset of the set of useless actions required to reach the first node.
2. A somewhat different approach has been explored in article by Hillas (*JET* 1994), the main difference being that whether or not beliefs are reasonable

should be judged relative to the set of 'reasonable' equilibria, not the set of all equilibria. The main features are described by the following quote:

"DEFINITION 4. Let $M \subset \Phi$. Say $(\mu, \pi) \in \Phi$ is *not believed in the first sense relative to* M if there exists $h \in H$ such that h is not reached with positive probability under (μ, π) but is reach with positive probability under some element of M and $\pi^{\iota(h)}(h)$ is not a best response to some beliefs in the convex hull of the beliefs at h of all elements of M that reach h with positive probability.

"Say $(\mu, \pi) \in \Phi$ is *not believed in the second sense relative to* M if there exists $h \in H$ such that h is not reached with positive probability under (μ, π) but is reach with positive probability under some element of M and $\mu|_h$ does not lie in the convex hull of the beliefs at h of all elements of M that reach h with positive probability.

"Say $(\mu, \pi) \in \Phi$ is *not believed in the third sense relative to* M if there exists $h \in H$ such that h is not reached with positive probability under (μ, π) but is reach with positive probability under some element of M and $\pi^{\iota(h)}(h)$ does not lie in the convex hull of the strategies played at h in those elements of M that reach h with positive probability.

"DEFINITION 5. A *solution* so an extensive game is a set $K \subset \Phi$. Say that the solution K is a *first (second) (third) order internally stable set of beliefs* if no element of K in not believed in the first (second) (third) sense relative to K. Say that the solution K is a *first (second) (third) order externally stable set of beliefs* if every element of $\Phi - K$ in not believed in the first (second) (third) sense relative to K. Call K a *first (second) (third) order stable set of beliefs* if K is both internally and externally stable."

a. There is a slight difference between Hillas' notion of beliefs and Kreps and Wilson's, which is not apparent here, but would show up if one considered these definitions carefully.

b. The general notions of internal and external stability of "standards of behavior" are due to Greenberg, and have been applied in a number of contexts.

3. Justifiability is not invariant with respect to transformations of the tree.

E. The relation between justifiability and stability

1. A stable (or hyperstable, essential, etc.) connected component of the set of agent normal form equilibria contains the strategy component of a justifiable equilibrium.

   a. This allows justifiability to be used in the computation of the stable equilibria and related concepts.

2. Stability is more restrictive than justifiability.

   a. Consider Figure 2.