87-F-3

Forward Induction and Equilibrium Refinement*

by

Masahiro Okuno-Fujiwara
University of Tokyo, and
University of Pennsylvania

and

Andrew Postlewaite
University of Pennsylvania

August, 1986

Revised February, 1987

PREPARED UNDER
THE PROJECT ON APPLIED ECONOMICS
RESEARCH INSTITUTE FOR THE JAPANESE ECONOMY#

---

* This work was begun while Postlewaite was visiting University of Tokyo under a grant by the Japan Society for the Promotion of Science. Their support and support from the National Science Foundation is gratefully acknowledged. We are grateful for comments made by Ken Binmore, George Mailath and Motty Perry. We thank Jeff Banks for pointing out an error in an example in an earlier version.

1. INTRODUCTION

There are many economic problems which, modelled by games of incomplete information using sequential equilibrium as the solution concept, give rise to many (often infinitely many) sequential equilibria. Often many of these equilibria seem implausible because of the beliefs associated with some disequilibrium information sets (i.e., disequilibrium beliefs.) Recently a number of papers have appeared proposing various refinements of the set of sequential equilibria based on tests of the plausibility of the disequilibrium beliefs (see, e.g., McClennan [1985], Kreps [1984], Cho and Kreps [1986], Banks and Sobel [1986], Cho [1986], Farrell [1985] and Grossman and Perry [1986]). Some of these papers are restricted to signalling games; this paper will also be restricted to such games.

A signalling game is a game in which there are only two players, I and II. Player I possesses private information, modelled by identifying a type for player I with each different information he might have. On the basis of his type, he sends a message to II. Player II, upon observing the message player I sends but without knowledge of the true type, then takes an action (reply). The payoffs to each player are determined by player I's true type (his private information), his message, and player II's action.

Many of the refinements mentioned above are based on the idea that some disequilibrium beliefs are implausible in that they put positive probability on some types of player I which are not likely to send this disequilibrium message. Kreps [1984] and Cho and Kreps [1986] refinement may be summarized as follows. Fix a sequential equilibrium and let m be an associated disequilibrium message (a message which no type of player I will send in the equilibrium.) Suppose there is a set of types (say K) who would never want to send this message because, regardless of the beliefs player II would form

upon observing this message and regardless of the optimal response II would subsequently choose, these types would obtain smaller payoffs than they would have obtained had the original equilibrium been played. Suppose further that for any beliefs that puts no weight on the set K of types which must be worse off by sending m, any best response by II yields a higher payoff than the original equilibrium payoff to all types not in K. Then the original equilibrium is said to fail their **intuitive criterion**. Clearly, the original equilibrium must have been sustained because disequilibrium beliefs associated with m are not consistent with the above observation. Specifically, the original disequilibrium belief must necessarily have put positive probability on (some subset of) K.

The idea behind their argument is typically illustrated by the following argument. A sequential equilibrium with a disequilibrium message satisfying the above property collapses when some type of player I plays this message with the following speech to player II. "I am sending disequilibrium message m and you should believe that I am of type t (or in some subset of types). If I were of other types, I would never have sent this message because no matter how you interpret this speech I would have obtained higher payoff had I not sent this message. Any interpretation of this message which excludes those other types which would not wish to send this message will lead to a higher payoff for the type t (or set of types) which I claim to be."

Grossman and Perry proposed a variant (further refinement) of Kreps and Cho's idea by insisting that all disequilibrium beliefs must be consistent with a similar line of reasoning. Their refinement can roughly be thought of as follows. Fix a sequential equilibrium and consider any disequilibrium message. For each subset K of types, we ask whether all members of K would obtain higher payoff (than the equilibrium payoff) if player II

would obtain higher payoff (than the equilibrium payoff) if player II conjectured that it is precisely these types who are sending this message and formed his belief and chose his optimal response accordingly. If there is a message m and a nonempty set K satisfying such a condition, we say that the original equilibrium fails the perfect sequentiality test. Grossman and Perry's refinement requires that the beliefs associated with each information set must be formed so that, when there is a set K of types which would benefit by playing the message, then player II must believe exactly the set K of types must be sending this message.

These refinements rely on the idea of **forward induction** (see Kohlberg and Mertens [1985]). In the words of Kohlberg and Mertens, a disequilibrium move and the resulting subgame should be considered as a "very specific form of pre-play communication." Any disequilibrium move should mean that the player is [effectively] sending the following message to other players. "Look, I had the opportunity to play the equilibrium strategy, and nevertheless I decided to play this move, and my move is already made. We both know that you can no longer talk to me, because we are in the game, and my move is made. So think now well, and make your decision."

In the usual interpretation of Nash equilibrium, the concept of forward induction may be justified as follows. A signalling game is to be played with the help of a computer. At the outset of the game, each player is asked to program his choice of move for each information set knowing that once he finishes his programming he can no longer change his choice. An outside referee suggests a sequential equilibrium to each player. In contemplating whether the suggested equilibrium is a compelling one, player II asks himself whether the beliefs associated with a disequilibrium information set is sufficiently plausible. In terms of Cho and Kreps argument, for example, he

may wonder whether some type of player I might deviate from the suggested equilibrium and actually choose the move that leads to the disequilibrium information set. He would find the deviation likely to happen if the equilibrium fails the Kreps-Cho intuitive criterion, for there is certainly an incentive for those types to deviate from the proposed equilibrium. Then player II would reject the proposed equilibrium because he now found a compelling argument, that the disequilibrium move will actually be made if player I happens to be a proper type, making the originally suggested "equilibrium" no longer a consistent proposal of plays.

If we interpret arguments of Kreps-Cho and Grossman-Perry this way, however, a problem emerges. Suppose player II is convinced that a set K of types would choose a disequilibrium message m. This implies that those types in K will no longer choose the original (suggested) equilibrium messages. Thus, player II's beliefs at the information sets associated with these messages must be revised, leading to different replies by player II. Since, by assumption, this argument is compelling enough to change II's beliefs, those types not in K should similarly find the same argument compelling. The change in player II's reply then may change the choice of messages by those types not in K, as they may find that choosing the original messages would yield a lower payoff than before. As a matter of fact, they might choose the message m instead. After these revisions the set of types who will choose the message m is no longer equal to K and it is no longer apparent that the original disequilibrium belief is implausible.

Thus, when player II is contemplating whether the suggested belief is appropriate at an information set, he must check whether his beliefs at all other information sets are consistent with his logic simultaneously. In other words, the "speeches" that player I attempts to convey to II with forward

induction must include not only what belief at the disequilibrium information set he proposed and what reply II should choose accordingly, but also how all other types might play and what beliefs II should form at all other information sets. For the speech to be compelling, all of these proposed strategy choices and beliefs must be consistent. That is, the speech must be a proposal of an alternative sequential equilibrium.

This argument does not preclude the possibility of using forward induction to refine the set of equilibria. It means that when player II contemplates the appropriateness of a sequential equilibrium and of the beliefs associated with a disequilibrium message m, he must check whether some types of player I might want to deviate from the suggested equilibrium by proposing an alternative equilibrium. Suppose there is an alternative sequential equilibrium for which the message m is played in the equilibrium by some set of types. If the set which plays m prefers the alternative equilibrium, the given sequential equilibrium should be rejected. We shall describe such a situation as the alternative equilibrium **defeating** the original equilibrium.

The plan of this paper is as follows. In section 2, we present the formal model and present examples showing how our refinement differs from those of Cho-Kreps and Grossman-Perry. In section 3 we prove that the set of (pure strategy) sequential equilibria which satisfy our refinement test is non-empty for an important class of signalling games. We prove this by showing a particular equilibrium always satisfies our refinement test. Section 4 contains concluding remarks including the relationship between our refinement and other refinement notions based upon perturbation.

2. FORWARD INDUCTION

2.1 PERFECTLY SEQUENTIAL EQUILIBRIUM

We begin with notation. There are two players, I and II.

$T = \{1,\ldots,n\}$	set of player I's types,
$p(t)$	probability of type t; assumed to be common knowledge,
M	set of moves for player I,
R	set of moves for player II,
$u(m,r,t)$	payoff for type t of player I for the pair of moves $(m,r) \in M \times R$,
$v(m,r,t)$	payoff for player II for the pair of moves $(m,r) \in M \times R$ when player I is of type t,
Δ_M, Δ_R, Δ_T	set of all probability distributions on M, R and T respectively,
$\mu: T \to \Delta_M$	a mixed strategy for I,
$\rho: M \to \Delta_R$	a mixed strategy for II,
$\beta: M \to \Delta_T$	II' belief function, assigning a probability distribution over T upon observing m
$\mu(m\|t)$	probability I plays m when his type is $t \in T$,
$\rho(r\|m)$	probability II plays r when he observes $m \in M$,
$\beta(t\|m)$	II's conditional belief over T when he observes $m \in M$,
$U(m,\rho(m),t)$	expected payoff of sending m for type t of player I when his true type is t and when II's strategy is ρ, i.e., $U(m,\rho(m),t) = \Sigma_{r\in R}\, \rho(r\|m)\, u(m,r,t)$

$BR: M \times \Delta_T \to R$ set of best responses to m given $\beta(m)$, i.e.

$BR(m,\beta(m)) = \arg\max_{r\in R} \Sigma_{t\in T}\, \beta(t|m)\, v(m,r,t)$.

Definition: $\sigma^* = (\mu^*,\rho^*,\beta^*)$ is a sequential equilibrium if:

1) $\forall\, m \in M,\ t \in T\ \ \mu^*(m|t) > 0$ only if $m \in \operatorname{argmax}_{m'\in M} U(m',\rho^*(m'),t)$,

2) $\forall\, r \in R,\ m \in M\ \ \rho^*(r|m) > 0$ only if $r \in BR(m,\beta^*(m))$,

3) $\forall\, t \in T$, and $m \in M\ \ \beta^*(t|m) = \dfrac{p(t)\,\mu^*(m|t)}{\sum_{t'\in T} p(t')\,\mu^*(m'|t)}$ if the denominator is positive.

With an abuse of notation we will write $u(\sigma^*,t)$ to be the expected payoff associated with σ^* for type t.

Grossman and Perry (1986) introduced a refinement of sequential equilibrium based on a restriction of the beliefs that an agent could hold at disequilibrium information sets. Roughly speaking, Grossman and Perry restrict an agent finding himself at an information set which should not have been reached during the play of the game to "try to interpret the move as a signal by the player I." They test a given sequential equilibrium in the following manner. For each information set which is not reached in the given equilibrium, player II hypothesizes that the move was made by some set of types of player I and revises his prior by Bayes rule conditional upon player I being in the specified set of types. If his best response given these beliefs is preferred by precisely the prespecified set of types, the given sequential equilibrium is said to fail the Grossman-Perry test.

Formally a sequential equilibrium $\sigma^* = (\mu^*, \rho^*, \beta^*)$ fails the perfect sequentially test if $\exists\ m \in M$, $\beta(m) \in \Delta_T$, $\rho(m) \in \Delta_R$ and $\pi : T \to [0,1]$ such that:

(1) $\forall t: \mu^*(m|t) = 0$,

(2) $\forall t: \beta(t|m) = \dfrac{p(t)\,\pi(t)}{\Sigma_{t' \in T}\, p(t')\pi(t')}$,

(3) $\forall r: \rho(r|m) > 0$ only if $r \in BR(m, \beta(m))$,

(4) $\pi(t) = 1$ if $u(\sigma^*, t) < U(m, \rho(m), t)$,

$\pi(t) = 0$ if $u(\sigma^*, t) > U(m, \rho(m), t)$,

(5) $\{t \mid u(\sigma^*, t) < U(m, \rho(m), t)\} \neq \phi$.

A sequential equilibrium which passes this test is said to be **perfectly sequential**.

Condition (1) states that m is a disequilibrium message. Conditions (2)-(5) state that, if player II conjectures that each type t of player I will play m with probability $\pi(t)$ and hence his posterior belief becomes $\beta(m)$ by Bayes rule (condition (2)), then $\rho(m)$ is his best response (condition (3)). Moreover, with such a conjecture, precisely the non-empty set described in (5) will become better off and all types in this set are conjectured to send this message with probability 1 (the first half of condition (4)), while no type whose payoff becomes lower is conjectured to send this message (latter half of condition (4)).

For some commonly analyzed games, the Grossman-Perry test rejects what seem to be unintuitive equilibria, leaving more reasonable equilibria. However it may be that **all** sequential equilibria fail this test. Grossman and Perry provide an example of such a game attributed to Maskin.

We provide below another example of a game in which the set of perfectly sequential equilibrium is empty. In this example it is somewhat easier to see why the set is empty than in other examples (we hope); also, the example illustrates how the concept may be altered in interesting ways.

Consider the game in figure I. Player 1 has four moves while player II has five. One equilibrium of this game is that player I plays strategy 4, giving both player I and II payoffs of 2 regardless of type. A set of disequilibrium beliefs for player II which support this as an equilibrium have player II believing that strategy i is played by type i of player I and playing strategy i as his best response. These beliefs do not satisfy the Grossman-Perry test. If player II observes strategy 1, he could conjecture that it was played by types 1 and 3, giving rise to a probability distribution over the types of $(1/2,0,1/2)$, that is, that is equally likely that the types 1 or 3 played this strategy but never type 2. With these beliefs, II's best response is to play strategy 4. The set of types who prefer this outcome to the proposed equilibrium consists of types 1 and 3, the two types with positive probability in the proposed beliefs. Type 2 would strictly prefer the outcome in the existing equilibrium to this outcome. It should be noted that only beliefs that lead to player II choosing strategy 4 make type 1 (or any other type in fact) better off than in the existing equilibrium. These beliefs, in the case that strategy 1 is played, then rule out the beliefs that supported the original equilibrium described above. Thus the proposed equilibrium is not perfectly sequential. It should be noted that the set

consisting of types 1 and 3 is the only set which could be "conjectured" by player II which leads to a rejection of the proposed equilibrium.

In a similar manner, player II could conjecture upon seeing strategy 2 that it was chosen by types 1 and 2 which would lead him to a probability distribution (1/2,1/2,0) in the case that strategy 2 was chosen by player I. With these beliefs his optimal choice is strategy 4, which is preferred to the existing equilibrium by types 1 and 2 but less preferred by type 3. Thus the beliefs in the proposed equilibrium which follow player I's choosing the non-equilibrium strategy 2 also fail the Grossman-Perry test. Similarly, the beliefs of player II given strategy 3 by player I fail the test since a conjecture that this strategy is played by types 2 and 3 leads to beliefs of (0,1/2,1/2) and a best response by II of strategy 4. This leads to an outcome preferred by types 2 and 3 and less preferred by 1.

Thus there are beliefs for player II following any disequilibrium move by player I such that II's best response with these beliefs is to play strategy 4 which in each case is preferred to the existing equilibrium outcome by precisely those types of player I which were conjectured to have played the disequilibrium strategy. None of the disequilibrium beliefs associated with any non-equilibrium move by player I in the proposed sequential equilibrium satisfy the Grossman-Perry test.

But now suppose that player II were to follow the suggested logic. If the logic is compelling, player I ought to be able to calculate in the same way as player II does. What would player I do? Since any disequilibrium move by player I leads player II to play strategy 4, type 1 should play strategy 1 because this leads to the highest possible utility. It is true that playing strategy 2 also leaves type 1 better off than in the existing equilibrium, but we should assume that he would optimize. Similarly, type 2 would play

strategy 2 and type 3 would play strategy 3. In summary, player II's original disequilibrium beliefs were ruled out by the Grossman-Perry test because of the existence of other "self-fulfilling" beliefs on II's part. But if II were to adopt these beliefs and behave optimally with these beliefs, the optimal behavior on I's part preceding II's move would not support these revised beliefs, but rather, would support II's original beliefs. Player II, who can calculate these changes in player I's choice of optimal strategies, might not be persuaded to change his original beliefs.

The argument above suggests that the mere existence of other "self-fulfilling" beliefs in the sense of Grossman-Perry is not enough to convince player II to change his beliefs, because changing his beliefs is bound to create further adjustments in player I's choice of strategies, which would force further revisions in player II's beliefs, creating yet further change in I's optimal strategies and so on. If both players I and II are rational and if it is common knowledge that they are rational, player II would change his disequilibrium beliefs only when all types of the prescribed set of player I types prefer to choose the strategy after taking account of all the subsequent adjustments that will take place once such revisions in disequilibrium beliefs are made. But once all the subsequent adjustments are made, we must be at an equilibrium; if not some further adjustments should be contemplated. Hence, if players are to engage in an exercise such as is involved in the definition of perfectly sequential equilibrium and all players carry the forward induction to completion, we are led to a test as follows. Consider a proposed sequential equilibrium. For each information set which is not to be reached in equilibrium, player II should conjecture that there is a set of types of player I which is playing an alternative equilibrium. If there is an alternative equilibrium for which some non-empty set of types of player I

choose the given strategy and that set is precisely the set of types who prefer the alternative equilibrium to the proposed equilibrium, forward induction requires that the beliefs associated with this information set in the original equilibrium be consistent with this set. If the beliefs are not consistent with forward induction, we say the second equilibrium <u>defeats</u> the proposed equilibrium.

Formally, we say that an equilibrium $\sigma = (\mu,\rho,\beta)$ defeats another equilibrium $\sigma' = (\mu',\rho',\beta')$ if $\exists\ m \in M$, and $\phi \neq K \subset T$ such that:

(1) $\forall t$: $\mu'(m|t) = 0$ and $K = \{t \in T \mid \mu(m|t) > 0\}$,

(2) $\exists t \in K$: $u(\sigma,t) > u(\sigma',t)$ and $\forall\ t \in K$: $u(\sigma,t) \geq u(\sigma',t)$,

(3) For all $\pi: T \to [0,1]$ and $q \in \Delta_T$ satisfying

(a) $\pi(t) = 1$ if $t \in K$ and $u(\sigma,t) > u(\sigma',t)$, and

$\pi(t) = 0$ if $t \notin K$,

(b) $$q(t) = \frac{p(t)\pi(t)\mu(m|t)}{\Sigma_{t' \in T}\, p(t')\pi(t')\mu(m|t')},$$

$\beta'(m) \neq q$.

Condition (1) says that m is a disequilibrium message at σ', but this message is sent with positive probability by a non-empty set of types K in σ. Condition (2) says that all members of K are as well off in σ as in σ' and some members are strictly better off. Condition (3) says that if those types who are strictly better off send this message with probability 1 and if those who are worse off send it with probability 0, the resulting beliefs (as defined in (3b)) are not equal to the original disequilibrium

beliefs $\beta'(m)$. Condition (1) may be weakened to "K is included in the set $\{t \in T \mid \mu(m|t) > 0\}$," namely there may be some types outside K who send the message m with positive probability. Such a weakening of the concept of defeat would cause an equilibrium to be more easily defeated and thus make the set of undefeated equilibria smaller. For example, in the now famed Kreps' quiche-beer example, the quiche-quiche equilibrium is defeated by the beer-beer equilibrium if we use the weaker notion but not the stronger notion.

In the example presented above, there are three other sequential equilibria besides the one in which all three types of player I play strategy 4. Equilibrium 2 has types 1 and 3 of player I playing strategy 1 and type 2 playing strategy 4. The beliefs associated with the strategies not played in equilibrium are that strategy i is played by type i, i = 2,3. There are two other similar equilibria, equilibrium 3 in which types 1 and 2 play strategy 2 while type 3 plays strategy 4 and equilibrium 4 in which types 2 and 3 play strategy 3 while type 1 plays strategy 4. The disequilibrium beliefs are similar to those in the equilibrium described just above. None of these equilibria passes the test we proposed above. Equilibrium 2 defeats equilibrium 3, equilibrium 3 defeats equilibrium 4, equilibrium 4 defeats equilibrium 2 and each of these defeats the equilibrium in which all types of player I play strategy 4.

We believe that conjectures about the types of player I who might have made disequilibrium moves should satisfy a complete forward induction test, that is that the conjectures should be consistent with equilibrium behavior as in the notion of defeat. For the above example, however, the notion of defeat leads to the same set of resulting allocations as the set of perfectly sequential equilibria: the empty set. To see that the two notions do lead to different conclusions in some games, consider example 2, shown in figure 2.

This game is essentially a signalling game in a slightly different form than is usual. Player I is of two possible types and has four pure strategies; player II has three pure strategies. ε is any number between 0 and 1. There are two equilibria for this game in which player I plays a pure strategy. In the first (separating) equilibrium, type 1 plays strategy 1 followed by player II playing strategy 3, and type 2 plays strategy 4 followed by player 2 playing strategy 1. The beliefs associated with any disequilibrium move is that it was made by type 1. Payoffs are 3.5 and 3 respectively for each type. The second (pooling) equilibrium has both types playing strategy 2 and player II playing strategy 2. Again, the disequilibrium beliefs are that any other strategy was played by type 1. The associated payoffs are (4.5, 4.5).

Besides these two equilibria which have player I playing a pure strategy, there is an additional equilibrium in which he plays a mixed strategy. This equilibrium has type 1 of player 1 playing strategies 1 and 3 with probabilities ν and $1-\nu$ respectively where $\nu = 2\varepsilon/(1+\varepsilon)$, and type 2 playing strategy 3 with probability 1. Player II plays strategies 1 and 2 with probabilities 2/3 and 1/3 respectively when player I plays strategy 3. Player II plays strategy 3 otherwise. Payoffs are 3.5 and 4.33 respectively for each type.

None of these equilibria is perfectly sequential. For the first equilibrium, suppose the second player conjectures that strategy 2 was played by both types. In this case the probabilities that he should use are the prior (.5,.5) and a best response is to play strategy 2. This gives rise to expected payoffs of 4.5 for each type as opposed to the payoffs to the two types of 3.5 and 3 respectively in the original equilibrium. Thus, the separating equilibrium is not perfectly sequential.

Now consider the second equilibrium, the pooling equilibrium. Suppose that the second player conjectures that strategy 3 is played by type 2. Player II's best response is then to play strategy 2. If type 2 of player I

did play strategy 3, this would be an improvement over the proposed equilibrium which gives him 4.5. If type 1 were to play strategy 3, this would result in a payoff of 4 also, but this would be worse than at the proposed equilibrium which gives him an expected payoff of 4.5. Thus with the conjecture that strategy 3 comes from type 2, player II's best response is such that only type 2 would be better off. Thus this equilibrium also fails to be perfectly sequential.

The mixed strategy equilibrium is not perfectly sequential since if player II conjectures that strategy 2 comes from both types his best response is the same as in the pooling equilibrium above which gives both types higher expected payoffs than in the mixed strategy equilibrium. Thus the set of perfectly sequential equilibria is empty.

We will turn now to the set of undefeated equilibria. The separating equilibrium is defeated by essentially the same argument as that used to show that it was not perfectly sequential. There is an alternative equilibrium in which strategy 2 (which is a disequilibrium move for the separating equilibrium) is played. It is played by both types and both types are better off at this alternative equilibrium than at the proposed separating equilibrium. In exactly the same way, the pooling equilibrium defeats the mixed equilibrium. Strategy 3 is again a disequilibrium strategy for the given equilibrium which is played by both types in the pooling equilibrium. Since the pooling equilibrium is better for both types of player I, it defeats the given equilibrium.

The pooling equilibrium is not defeated however. It is clear that the pooling equilibrium is not defeated by any equilibrium since no player is better off in other equilibria than at the pooling equilibrium. Thus, the set of undefeated equilibrium is non-empty and contains only the pooling equilibrium.

The example above illustrates nicely the difference between the two

concepts. The pooling equilibrium is not perfectly sequential; the beliefs at the disequilibrium strategy 3 upset the equilibrium. The conjecture that this strategy was played by type 2 gives rise to a best response by player II which is preferred to the original equilibrium payoff by type 2 and not by type 1. But this conjecture is inconsistent with any equilibrium so it is not part of the test involved in the test to see whether this equilibrium is defeated. Thus the pooling equilibrium passes the test posed by the notion of defeat, but not that posed by perfect sequentiality.

2.2 THE INTUITIVE CRITERION

Grossman and Perry's use of a forward induction based refinement of sequential equilibria followed earlier work in a similar spirit by Kreps [1985] (see also Cho and Kreps [1986].) Kreps proposed a refinement based on what he called the intuitive criterion. It is useful to compare our notion of undefeated equilibrium with Krep's intuitive criterion. With an abuse of notation we write; $\forall\ S \subset T$ and $\forall\ m \in M$ $BR(m,S) = \bigcup_{\beta(m)\in\Delta_S} BR(m,\beta(m))$.

The Intuitive Criterion: Given a sequential equilibrium σ^*, for each disequilibrium message m, form the set $S(m)$ consisting of all types t such that

$$u(\sigma^*,t) > \max_{r\in BR(m,T)} u(m,r,t)$$

If for any message m there is some type $t' \in T$ (necessarily not in $S(m)$) such that

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

then the equilibrium is said to fail the Intuitive Criterion.

This refinement has a nice property not shared by either of the other two refinements that we have discussed so far, namely that the set of sequential equilibria satisfying the intuitive criterion is always nonempty. In the case that the set of undefeated equilibrium is nonempty, it may not be contained in the set of equilibria which satisfy the intuitive criterion; the sets may in fact be disjoint.

If we return to example 2 above, we see that the pooling equilibrium fails the intuitive criterion for roughly the same reason that it fails to be perfectly sequential. For any beliefs held by player II following strategy 3, type 1 cannot be better off than at the pooling equilibrium. Restricting beliefs to those putting probability 0 on the strategy having been played by type 1 leads to player II's choosing strategy 2. This outcome is preferred to the outcome in the pooling equilibrium by precisely type 2; thus, the pooling equilibrium fails the intuitive test. The mixed strategy equilibrium and the separating equilibrium, on the other hand, pass the intuitive criterion. The set of equilibria which satisfy the intuitive criterion is exactly the set of equilibria which is defeated. Banks and Sobel [1986] proposed a further refinement of Kreps intuitive criterion. Our refinement differs from that of Banks and Sobel for this example as well.

3. UNDEFEATED EQUILIBRIA IN A SIGNALLING GAME

3.1 MODEL AND DEFINITIONS.

In this section, we shall show that, in an economically important class of (continuous strategy set) signalling games, the set of undefeated equilibria is non-empty. We shall show this by demonstrating a particular sequential equilibrium is always undefeated, namely the lexicographically

equilibrium is always undefeated. The following class of signalling games is economically important as many asymmetric information situations analyzed in the literature, such as the signaling game of Spence, the insurance model of Rothschild and Stiglitz, limit pricing model (e.g., Milgrom and Roberts [1982]) and some litigation models (e.g., Banks and Sobel [1986]), are all special cases of such games.

Consider a signalling game G where both M and R are each subsets of the non-negative half-line (and hence a continuum.) We shall confine our attention to the set of pure strategy equilibria. Thus in this section, strategies μ and ρ are mappings from T to M and from M to R, respectively. We denote the set of pure strategy sequential equilibria for the game G by $PSE(G)$.

$\sigma \equiv (\mu,\rho,\beta) \in PSE(G)$ <u>defeats</u> $\sigma' \equiv (\mu',\rho',\beta') \in PSE(G)$ if $\exists\ m \in M$ and $\phi \neq K \subset T$ such that:

(1) $\forall t: \mu'(t) \neq m$, and $K = \{t \in T | \mu(t) = m\}$

(2) $\forall t \in K: u(\sigma,t) \geq u(\sigma',t)$, and
$\exists t \in K: u(\sigma,t) > u(\sigma',t)$.

(3) For all $\pi: T \to [0,1]$ and $q \in \Delta_T$ satisfying

(a) $\pi(t) = 1$ if $t \in K$ and $u(\sigma,t) > u(\sigma',t)$, and
$\pi(t) = 0$ if $t \notin K$, and

(b) $q(t) = \dfrac{p(t)\,\pi(t)}{\sum_{t' \in T} p(t')\pi(t')}$,

$\beta'(m) \neq q$.

$\sigma \in PSE(G)$ is undefeated if there does not exist $\sigma' \in PSE(G)$ that defeats σ. $\sigma \in PSE(G)$ lexicographically dominates (ℓ-dominates) $\sigma' \in PSE(G)$ if there exists $j \in T$ such that $u(\sigma,j) > u(\sigma',j)$ and for $t > j$ $u(\sigma,t) \geq u(\sigma',t)$. $\sigma \in PSE(G)$ is the lexicographically maximum sequential equilibrium (LMSE) if there does not exist $\sigma' \in PSE(G)$ that ℓ-dominates σ. Finally, $\sigma \in PSE(G)$ is a completely separating equilibrium if $\forall t, t' \in T$, $\mu(t) \neq \mu(t')$ whenever $t \neq t'$.

We shall confine our attention to a class of signalling games which satisfy the following four assumptions.

Assumption 1: (Continuity and Concavity)

(i) M and R are closed, convex subsets of $\mathbb{R}^1$.

(ii) u and v are continuous in m and r.

(iii) v is strictly concave in r.

Remark: By A.1 (ii)-(iii), for any $q \in \Delta_T$ BR(m,q) is a single-valued continuous function on M. Hence u(m,BR(m,q),t) is well-defined.

Assumption 2: (Stochastic dominance)

$\forall t \in T$, $\forall m \in M$, $\forall q, q' \in \Delta_T$, whenever q stochastically dominates q', i.e., $\Sigma_{t' \leq t} q'(t') \geq \Sigma_{t' \leq t} q(t')$ for all $t \in T$ and strict inequality holds for some $t \in T$, $u(m,BR(m,q),t) > u(m,BR(m,q'),t)$.

Assumption 3: (Weak (resp. Strong) Monotonicity)

$\forall m, m' \in M$, $\forall r, r' \in R$, $\forall t, t' \in T$, if

(i) $u(m,r,t) \geq u(m',r',t)$,

(ii) $m \geq m'$ ($m > m'$, resp.), and

(iii) $t' > t$, then

(iv) $u(m,r,t') \geq u(m',r',t')$ $(u(m,r,t') > u(m',r',t'),$ resp.).

For the next assumption, we need an additional definition. For any nonempty subset K of T, $q_K \in \Delta_T$ is called <u>K-conditional belief</u> if:

$$q_K(t) = p(t)/\Sigma_{t' \in K} p(t') \qquad \text{if } t \in K$$

$$= 0 \qquad \text{otherwise.}$$

With an abuse of notation, we sometimes write the best response against m, when the belief is K-conditional belief, by $BR(m,K)$, i.e.,

for all $m \in M$ and for all subset K of T, $BR(m,K) = BR(m,q_K)$.

<u>Assumption 4</u>: (Satiation)

For all $t \in T$, all $m \in M$, and all $q \in \Delta_T$, if q is not the $\{n\}$-conditional belief and if $t \neq n$, then there exists $\bar{m}(m,BR(m,q),t) \in M$ such that for all $m' \geq \bar{m}(m,BR(m,q),t)$, $u(m',BR(m',\{n\}),t) < u(m,BR(m,q),t)$.

Assumption 1 is primarily a technical assumption. The second assumption states that all types of Player I prefer the (best) response of player II when player II believes I more likely to be of higher type. The third assumption is similar to a "single-crossing" property. It says that if some type t prefers a message-response pair (m,r) to a second pair (m',r'), when m is greater than m', then any type higher than t will also prefer (m,r) to (m',r'). This is to capture the idea that higher messages are "easier" for higher types to send than for lower types. The last assumption, 4, says sending very high message is prohibitively costly in the sense that there is a

message level such that no type, except possibly the highest type, would want to exceed even if the result was the most favorable possible beliefs on player II's part.

Main Results

Theorem 1. Under A1-A4, the LMSE is undefeated.

Theorem 2. Under A1-A4, if the LMSE is completely separating, it is the only undefeated pure strategy sequential equilibrium.

Remark: Thus, if the LMSE is pooling, there may be multiple undefeated pure strategy sequential equilibria. (Also see example 2 of the previous section for multiple mixed strategy undefeated equilibria.)

3.2 PROOFS

Lemma 1: If PSE(G) is non-empty, there exists a LMSE.

Proof: Trivial. Q.E.D.

Lemma 2: Under weak monotonicity, $\forall$ m, m' $\in$ M, $\forall$ r, r' $\in$ R, $\forall$ t, t' $\in$ T, if

(i) $u(m,r,t) \geq u(m',r',t)$

(ii) $u(m',r',t') > u(m,r,t')$, and

(iii) $t' > t$, then

(iv) $m' > m$.

Proof: Suppose, contrary to the assertion, $m \geq m'$. Then by weak monotonicity, (i) and (iii) imply

$$u(m,r,t') \geq u(m',r',t')$$

contrary to (ii). Q.E.D.

Next we define games truncated from G through restricting player I's types to be a subset of the original set, T. Formally for any $j \in T$, let

$$T^j = \{1,\ldots,j\}, \text{ and}$$

$$p^j(t) = q_{T^j}$$

A truncated game G^j is defined by substituting the subset T^j for T and the T^j-conditional belief p^j for p in the original game G. We shall denote the set of pure strategy sequential equilibria of G^j by $PSE(G^j)$.

The following property is important. Given any pure strategy equilibrium of the original game, $\sigma \in PSE(G)$, we define j-truncated equilibrium σ^j for $j \in T$ by simply deleting those types higher than j. It follows trivially that, as long as no type higher than j is sending the same message that j sends at σ, a j-truncated equilibrium σ^j is a pure strategy equilibrium in G^j.

Corollary to Lemma 2: If $\sigma \equiv (\mu,\rho,\beta) \in PSE(G^j)$, then $\mu(t) \leq \mu(t')$ whenever $t \leq t' \leq j$.

Proof: Trivial application of Lemma 2. Q.E.D.

Lemma 3: (Strong monotonicity implies reverse monotonicity) Under strong monotonicity, $\forall\, m, m' \in M$, $\forall\, r, r' \in R$, $\forall\, t, t' \in T$, if

(i) $u(m,r,t) \geq u(m',r',t)$,

(ii) $m < m'$, and

(iii) $t' < t$, then

(iv) $u(m,r,t') > u(m',r',t')$.

Proof: Suppose, contrary to the assertion, $u(m,r,t') \leq u(m',r',t')$. Then, by strong monotonicity, (ii) and (iii) imply $u(m',r',t) > u(m,r,t)$, contrary to (i). Q.E.D.

The next lemma is the key lemma for our proof. In effect, we shall prove the following. Suppose we are given two pure strategy equilibria of the original game, σ and $\hat{\sigma}$. If for some $j \in T$ j's equilibrium payoff is no smaller at $\hat{\sigma}$ than at σ, then we can construct yet another pure strategy equilibrium in $j+1$ truncated game in which every type less than or equal to j obtains a payoff at least equal to the payoff obtained at $\hat{\sigma}$ while $j+1$ obtains a payoff at least equal to the payoff obtained at σ.

Lemma 4: Suppose $\sigma \in PSE(G)$ and $\hat{\sigma} \in PSE(G^j)$ for some $j < n$. Suppose further that $u(\hat{\sigma},j) \geq u(\sigma,j)$. Let $H = \{t \mid t \geq j+1 \text{ and } \mu(t) = \mu(j+1)\}$ and $h = \max\{t \mid t \in H\}$. Then there exists $\sigma^* \in PSE(G^h)$ such that:

(i) $u(\sigma^*,t) \geq u(\hat{\sigma},t)$ for all $t \leq j$, and

(ii) $u(\sigma^*,t) \geq u(\sigma,t)$ for all $t \in H$.

Proof: The proof is by construction. We shall write for all t: $\mu(t) = m_t$, $\rho(\mu(t)) = r_t$, $\hat{\mu}(t) = \hat{m}_t$, and $\hat{\rho}(\hat{\mu}(t)) = \hat{r}(t)$.

Case 1: $u(\hat{m}_j, \hat{r}_j, j+1) \geq u(m_{j+1}, r_{j+1}, j+1) := u(\sigma, j+1)$.

We shall only prove that there exists $\sigma^* \in PSE(G^{j+1})$ which satisfies (i) for all $t \leq j$ and (ii) for $t = j+1$. For if this were the case, we can replace j in the original proposition by $j+1$ and $\hat{\sigma}$ by this new σ^*, and induction will prove the lemma.

Let $K = \{t | t \leq j \text{ and } \hat{m}_t = \hat{m}_j\}$ and let $k = \min\{t | t \in K\}$. By definition, $\hat{m}_j = \hat{m}_k$ and $\hat{r}_j = \hat{r}_k = BR(\hat{m}_j, K)$. Define:

$$m_K^* = \max \{m \in M | u(m, BR(m, K \cup \{j+1\}), k) \geq u(\hat{\sigma}, k)\}, \text{ and}$$

$$r_K^* = BR(m_K^*, K \cup \{j+1\}).$$

We must show that m_K^* exists. For this, denote the set defining m_K^* by M_K. We shall show that M_K is non-empty and bounded from above.

Non-emptiness follows because $\hat{m}_j$ is in M. To see this, observe that assumption 2 implies:

$$u(\hat{m}_j, BR(\hat{m}_j, K \cup \{j+1\}), k) > u(\hat{\sigma}, k), \tag{1}$$

for $K \cup \{j+1\}$-conditional belief stochastically dominates K-conditional belief.

M is bounded from above because, for any $m \geq \bar{m}(\hat{m}_j, \hat{r}_j, k)$

$$u(m, BR(m, K \cup \{j+1\}), k) < u(m, BR(m, \{n\}), k) < u(\hat{\sigma}, k)$$

always holds as the first inequality follows from the fact that the $\{n\}$-conditional belief stochastically dominates the $K \cup \{j+1\}$-conditional belief, while the second inequality holds from assumption 4.

Since $\hat{m}_j$ is in M_K, by (1);

$$\hat{m}_j < m_K^*. \tag{2}$$

Moreover, in view the definition of m_K^*;

$$u(m_K^*, BR(m_K^*, K \cup \{j+1\}), k) = u(m_K^*, r_K^*, k) = u(\hat{\sigma}, k). \tag{3}$$

We now define $\sigma^* := (\mu^*, \rho^*, \beta^*)$ in G^{j+1}. Let:

(a) for all $t < k$: $\mu^*(t) = \hat{\mu}(t)$,

(b) for all $t \in K \cup \{j+1\}$: $\mu^*(t) = m_k^*$,

(c) for all $m \leq \hat{m}_j$: $\beta^*(m) = \hat{\beta}(m)$ and $\rho^*(m) = \hat{\rho}(m)$,

(d) for all $m > \hat{m}_j$ with $m = m_K^*$: $\beta^*(m) = q_{K \cup \{j+1\}}$ and $\rho^*(m) = r_k^*$.

(e) for all $m > \hat{m}_j$ but $m \neq m_K^*$: $\beta^*(m) = q_{\{1\}}$ and $\rho^*(m) = BR(m, \{1\})$,

Namely, we preserve all equilibrium messages (a) and replies (c) associated with $\hat{\sigma}$ for the types smaller than k and messages no larger than $\hat{m}_j$. For those types in $K \cup \{j+1\}$, we assign the message m_K^* (b) and the associated reply r_K^* (d). For other messages, we assign the worst possible belief and the associated replies (e).

Clearly, the assertion (i) holds for all $t < k$. For $t \in K$, (i) follows because of monotonicity and (3). Finally, for $t = j+1$.

$$u(\sigma^*,j+1) = u(m_K^*,r_K^*,j+1) \geq u(\hat{m}_K,\hat{r}_K,j+1) = u(\hat{m}_j,\hat{r}_j,j+1) \geq u(\sigma,j+1)$$

where the first inequality follows from monotonicity and (3), and the last from the condition defining this case. Thus (ii) holds.

It remains to be shown that σ^* is indeed a sequential equilibrium of G^{j+1}. We must prove the following three properties:

(A) For any equilibrium message $m \in M$, i.e., $m = \mu^*(t)$ for some $t \in T$, $\beta^*(m)$ is formed by Bayesian updating;

(B) For any $m \in M$, $\rho^*(m) = BR(m,\beta^*(m))$;

(C) For any $t \in T$ and $m \in M$, $u(m_t^*, r_t^*, t) \geq u(m,\rho^*(m), t)$.

(A) is straightforward from the definition of conditional beliefs. (B) is straightforward as well. To prove (C), classify the following three subcases;

(C1) For $t < k$:

If m satisfies (c) in the definition of σ^*, then the fact that $\hat{\sigma}$ is itself a sequential equilibrium implies (C).

If m satisfies (e),

$$u(m_t^*,r_t^*,t) := u(\hat{\sigma},t) \geq u(m,\ BR(m,\hat{\beta}(m)),t)$$

$$\geq u(m,\ BR(m,\{1\}),t) := u(m,\rho^*(m),t)$$

where the first inequality holds from the fact that $\hat{\sigma}$ is a sequential equilibrium and the second from the fact that $\hat{\beta}(m)$ (weakly) stochastically dominates $\{1\}$-conditional belief.

If $m = m_K^*$,

$$u(m_t^*, r_t^*, t) := u(\hat{m}_t, \hat{r}_t, t) \geq u(\hat{m}_K, \hat{r}_K, t) > u(m_K^*, r_K^*, t)$$

as the first inequality follows from the fact that $\hat{\sigma}$ is itself a sequential equilibrium and the second inequality from (2), (3) and Lemma 3.

(C2) For $t \in K$:

If m satisfies (c),

$$u(m_t^*, r_t^*, t) = u(m_K^*, r_K^*, t) \geq u(\hat{\sigma}_t, t) \geq u(m, \hat{\rho}(m), t) := u(m, \rho^*(m), t)$$

where the two equalities follow from the definition of σ^* and the first inequality follows from (i). The second inequality follows from the equilibrium condition of $\hat{\sigma}$.

If m satisfies (e),

$$u(m^*, r^*, t) \geq u(\hat{\sigma}, t) \geq u(m, BR(m, \hat{\beta}(m)), t)$$

$$\geq u(m, BR(m, \{1\}), t) = u(m, \rho^*(m), t)$$

where the first inequality follows from (i), the second from the equilibrium condition $\hat{\sigma}$, the third from the fact that $\hat{\beta}(m)$ (weakly) stochastically dominates $q_{\{1\}}$.

(C3) for $t = j+1$:

For all $m \leq m_K^*$, monotonicity and results in (C2) implies (C).

For all $m > m_K^*$, (C) follows because the associated belief is the worst possible belief.

This proves the lemma for case 1.

Case 2: $u(\hat{m}_j, \hat{r}_j, j+1) < u(\sigma, j+1) := u(m_{j+1}, r_{j+1}, j+1)$.

By hypothesis, $u(\hat{\sigma}, j) \geq u(\sigma, j)$, i.e., $u(\hat{m}_j, \hat{r}_j, j) \geq u(m_j, r_j, j)$; combining this inequality with the equilibrium condition for σ, we obtain

$$u(\hat{m}_j, \hat{r}_j, j) \geq u(m_{j+1}, r_{j+1}, j).$$

In view of this inequality and the inequality characterizing the current case, lemma 2 implies $m_{j+1} > \hat{m}_j$.

Let $r_H^* = BR(m_H^*, H)$. Define:

$$m_H^* = \max \{m \in M \mid u(m, BR(m,H), j+1) \geq u(\sigma, j+1)\}.$$

By the definition of m_H^*, it follows that:

$$u(m_H^*, r_H^*, j+1) = u(m_{j+1}, r_{j+1}, j+1) = u(\sigma, j+1). \tag{4}$$

As in the previous case, the set defining m_H^* is non-empty and compact as m_{j+1} is in the set and it is bounded from above. Hence m_H^* is well defined.

Define $\sigma^* := (\mu^*, \rho^*, \beta^*)$ in the following way:

(a) for all $t \leq j$: $\mu^*(t) = \hat{m}_t$,

(b) for all $t \in H$: $\mu^*(t) = m_H^*$,

(c) for all $m \leq \hat{m}_j$: $\beta^*(m) = \hat{\beta}(m)$ and $\rho^*(m) = \hat{\rho}(m)$,

(d) for all $m > \hat{m}_j$ with $m = \hat{m}_H$: $\beta^*(m) = \Sigma_H$ and $\rho^*(m) = r_H^*$.

(e) for all $m > \hat{m}_j$ but $m \neq m_H^*$: $\beta^*(m) = q_{\{1\}}$ and $\rho^*(m) = BR(m, \{1\})$,

Property (i) follows trivially and (ii) follows from (4) for $j+1$ and from (4) and monotonicity for other $t \in H$. Using essentially the same arguments as in Case 1, σ^* is readily established as a sequential equilibrium of G^h. Q.E.D.

Proof of Theorem 1:

Let σ^* be the LMSE and suppose, contrary to the supposition, there exists $\sigma \in PSE(G)$ that defeats σ^* via (m,J).

Let $j = \max\{t|t \in J\}$, $k = \min\{t|t \in J\}$, $H = \{t \geq j+1 | m_t^* = m_{j+1}^*\}$ and $h = \max\{t|t \in H\}$. The restriction of σ on $T^j = \{1,2,\ldots,j\}$ is obviously a PSE of G^k and $u(\sigma,t) \geq u(\sigma^*,t)$ for all $t \in J$ with strict inequality for at least one $t \in J$ by the definition of defeat. Thus, by the previous lemma, there exists $\sigma' \in PSE(G^h)$ such that:

$$\text{for all } t \leq j: \quad u(\sigma',t) \geq u(\sigma,t), \text{ and}$$

$$\text{for all } t \in H: \quad u(\sigma',t) \geq u(\sigma^*,t).$$

It follows that:

$$\text{for all } t \in J \cup H: \quad u(\sigma',t) \geq u(\sigma^*,t), \text{ and}$$

$$\text{for some } t \in J: \quad u(\sigma',t) > u(\sigma^*,t).$$

But then, we can apply lemma 4 again, and there exists $\sigma'' \in PSE(G^{h'})$, where $h' = \max\{t|t \in H'\}$ and $H' = \{t|t \geq h+1 \text{ and } m_t^* = m_{h+1}^*\}$, satisfying:

for all $t \leq h$: $u(\sigma'',t) \geq u(\sigma',t)$, and

for all $t \in H'$: $u(\sigma'',t) \geq u(\sigma^*,t)$.

It then follows that:

for all $t \in J \cup H \cup H'$: $u(\sigma'',t) \geq u(\sigma^*,t)$, and

for some $t \in J$: $u(\sigma'',t) > u(\sigma^*,t)$.

Repeating the argument, we obtain $\hat{\sigma} \in PSE(G)$, a pure strategy equilibrium for the original game, satisfying:

for all $t \geq k$: $u(\hat{\sigma},t) \geq u(\sigma^*,t)$, and

for some $t \in J$: $u(\hat{\sigma},t) > u(\sigma^*,t)$.

But then $\hat{\sigma}$ ℓ-dominates σ^*, contrary to our supposition that σ^* is the LMSE. Q.E.D.

Proof of Theorem 2:

In view of theorem 1, we only have to show that, if the LMSE is a separating equilibrium, there is no undefeated pure strategy equilibrium which is not the LMSE. So suppose, contrary to our assertion, there exists an undefeated $\sigma \in PSE(G)$ which is ℓ-dominated by the LMSE, σ^*. Then there exists $j \in T$ such that:

(1) $u(\sigma^*,j) > u(\sigma,j)$,

(2) $u(\sigma^*,t) \geq u(\sigma,t)$ for all $t > j$.

We shall assume that $\mu^*(j)$ is a disequilibrium message under σ, for

otherwise we can construct $\sigma^{**} \in PSE(G)$ slightly perturbed from σ^* by assigning a new message $\mu^{**}(j)$ to j from the open interval $(\mu^*(j), \mu^*(j+1))$ and a new reply $\rho^{**}(\mu^{**}(j)) \in BR(\mu^{**}(j),\{j\})$, still preserving

(i) $u(\sigma^{**},j) > u(\sigma,j)$, and

(ii) $u(\sigma^{**},t) \geq u(\sigma,t)$ for all $t > j$.

Note that we can construct the above perturbed equilibrium because the LMSE, σ^*, is separating equilibrium. Otherwise, a change in message $\mu^*(j)$ may affect the equilibrium payoff of other types, in particular type $j+1$.

Obviously, $\beta(\mu^*(j))$ is not $q_{\{j\}}$, for otherwise an equilibrium condition is violated. Since $\mu^*(j)$ is sent only by type j as σ^* is separating equilibrium, σ^* defeats σ via $(\mu^*(j),\{j\})$, contrary to our suposition. Q.E.D.

4. CONCLUDING REMARKS

The definition of defeat might be altered in several interesting ways. In the test of whether an equilibrium is defeated, each disequilibrium strategy is conjectured to be a "signal" by some set of types that an alternative equilibrium is being played. If there are several alternative equilibria in which a given disequilibrium message is sent, player II will not be able to make an unambiguous comparison between the proposed equilibrium and possible alternatives. The test of whether a given equilibrium is defeated or not could be strengthened to ask whether there is a **unique** alternative equilibrium in which the given disequilibrium message is played and which has the additional properties in the definition of defeat. This would make it

less likely that an equilibrium is defeated and would make the set of undefeated equilibria larger. It would not solve the problem that for general games the set of undefeated equilibria may be empty, but it might make the concept of defeat more plausible.

In the definition of defeat, an equilibrium is tested according to the beliefs held by player II at disequilibrium information sets. In this sense the refinement is similar to that introduced by Grossman and Perry. The test to which an equilibrium is put to determine whether it is defeated or not is different in that it may be used to generate a partial ordering on the set of sequential equilibria. In the case that the set of undefeated equilibrium is empty, we can form a set of sequential equilibria such that every sequential equilibrium is defeated by a sequential equilibrium in this set. A minimal set, that is, a set such that no proper subset has the property must trivially exist. There may be multiple sets, but such a set would constitute a refinement which trivially must be non-empty. For some examples, this may reduce the set of equilibria in reasonable ways.

A third way in which the concept might be altered is to extend the definition to more general games. It should be straightforward to define the concept of defeat for games with more stages and with more than two players. For each history which is an equilibrium history except for a single move by some agent, we can ask whether there is an alternative equilibrium which is consistent with the history. If there is we can again compare the payoffs to the types who prefer the alternative equilibrium to the proposed equilibrium to determine whether the disequilibrium message can be interpreted as a signal that the alternative equilibrium is being played by that set of types.

In this paper, we have taken the position that refinement may be carried out in terms of forward induction. Alternatively, we can choose a refinement

based upon a perturbation of the equilibrium strategy. Many refinements, such as Selten [1975], Myerson [1978] and Kalai and Samet [1984], use such perturbations to refine Nash equilibria. Kohlberg and Mertens [1985] base their concept of stability upon a similar consideration as well. Perturbation based refinements and forward induction based refinements are not unrelated, however. As is evident from the discussion of Cho and Kreps [1985] and Banks and Sobel [1986], the two types of refinements are inherently related. Our concept of defeat can be alternatively defined as follows in terms of perturbations (generically at least).

Suppose there are two sequential equilibria σ and σ'. We write σ_i and σ'_i to denote corresponding behavioral strategy of player $i \in N$ where N is the set of players. Suppose for any $\varepsilon > 0$, there exists some δ_0 such that for any $\{\delta_i\}_{i\in N}(0 < \delta_i < \delta_0)$, the perturbed game where every strategy of player i is replaced by $(1-\delta_i)s + \delta_i\sigma_i$ has no equilibrium ε-close to the equilibrium σ'. Then, we could say σ **strongly defeats** σ'. It can be trivially seen that concept of **defeat** and that of **strong defeat** are generically equivalent.

Bibliography

Banks, J.S. and J. Sobel (1986), "Equilibrium Selection in Signalling Games," mimeo, forthcoming in Econometrica.

Cho, In-Koo and D. Kreps (1986), "More Signalling Games and Stable Equilibria," mimeo.

Cho, In-Koo (1986), "A Refinement of the Sequential Equilibrium Concept," mimeo.

Farrell, J. (1985), "Credible Neologisms in Games of Communication," mimeo.

Grossman, S.J. and M. Perry (1986), "Perfect Sequential Equilibrium," Journal of Economic Theory 39, 97-119.

Kohlberg, E. and J. Mertens (1985), "On the Strategic Stability of Equilibria," mimeo, CORE.

Kreps, D. (1984), "Signalling Games and Stable Equilibria," mimeo.

Kreps, D. and R. Wilson (1982), "Sequential Equilibria," Econometrica 50, 863-894.

McClennan, A. (1985), "Justifiable Beliefs in Sequential Equilibrium," Econometrica 53, 889-904.

Milgrom, P. and J. Roberts (1982), "Limit Pricing and Entry under Incomplete Information: An Equilibrium Analysis," Econometrica 50, 443-459.

Rothschild, M. and J.E. Stiglitz (1976), "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information," Quarterly Journal of Economics, 80, 629-49.

Spence, M. (1974), Market Signalling, Harvard University Press, Cambridge.

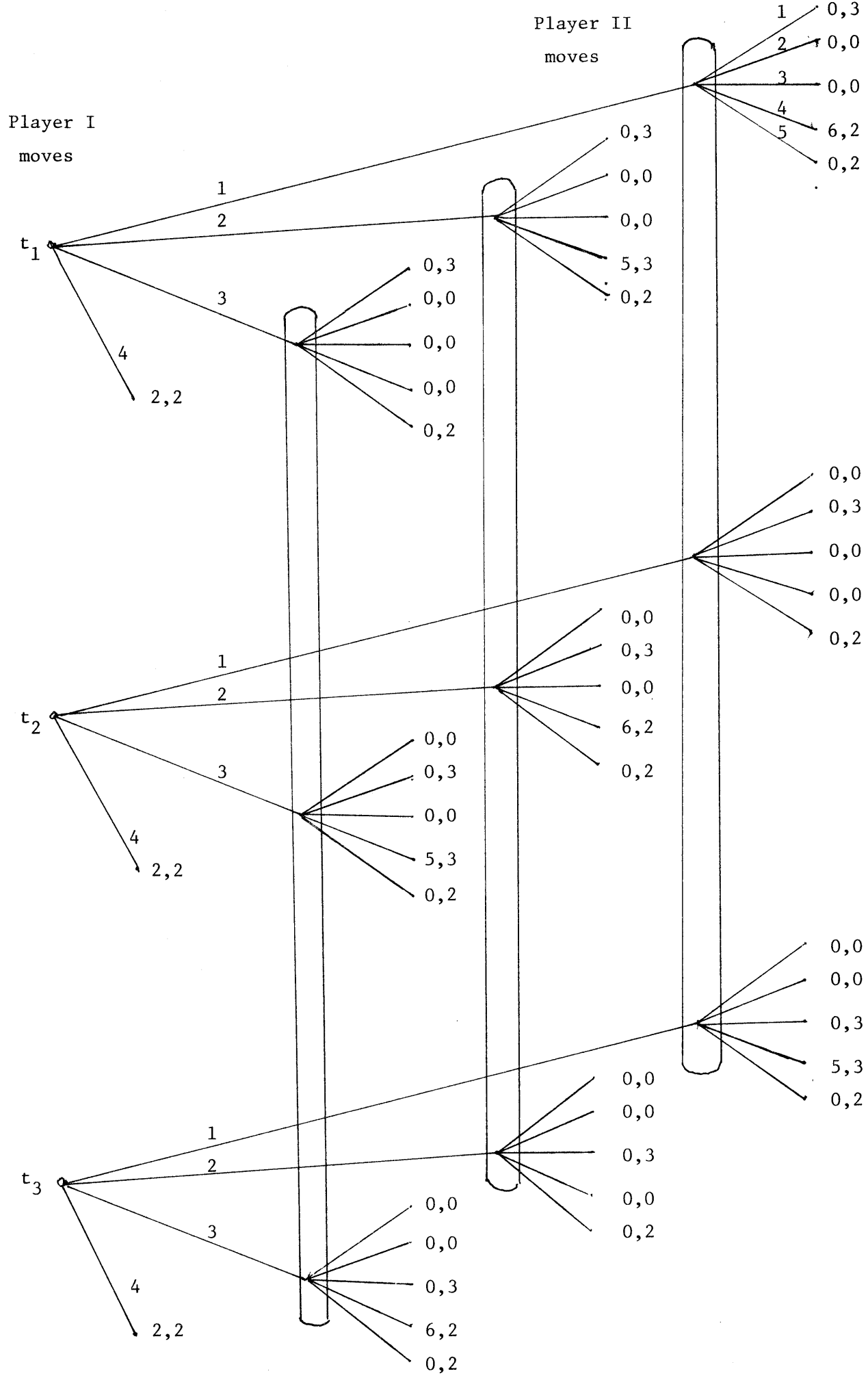
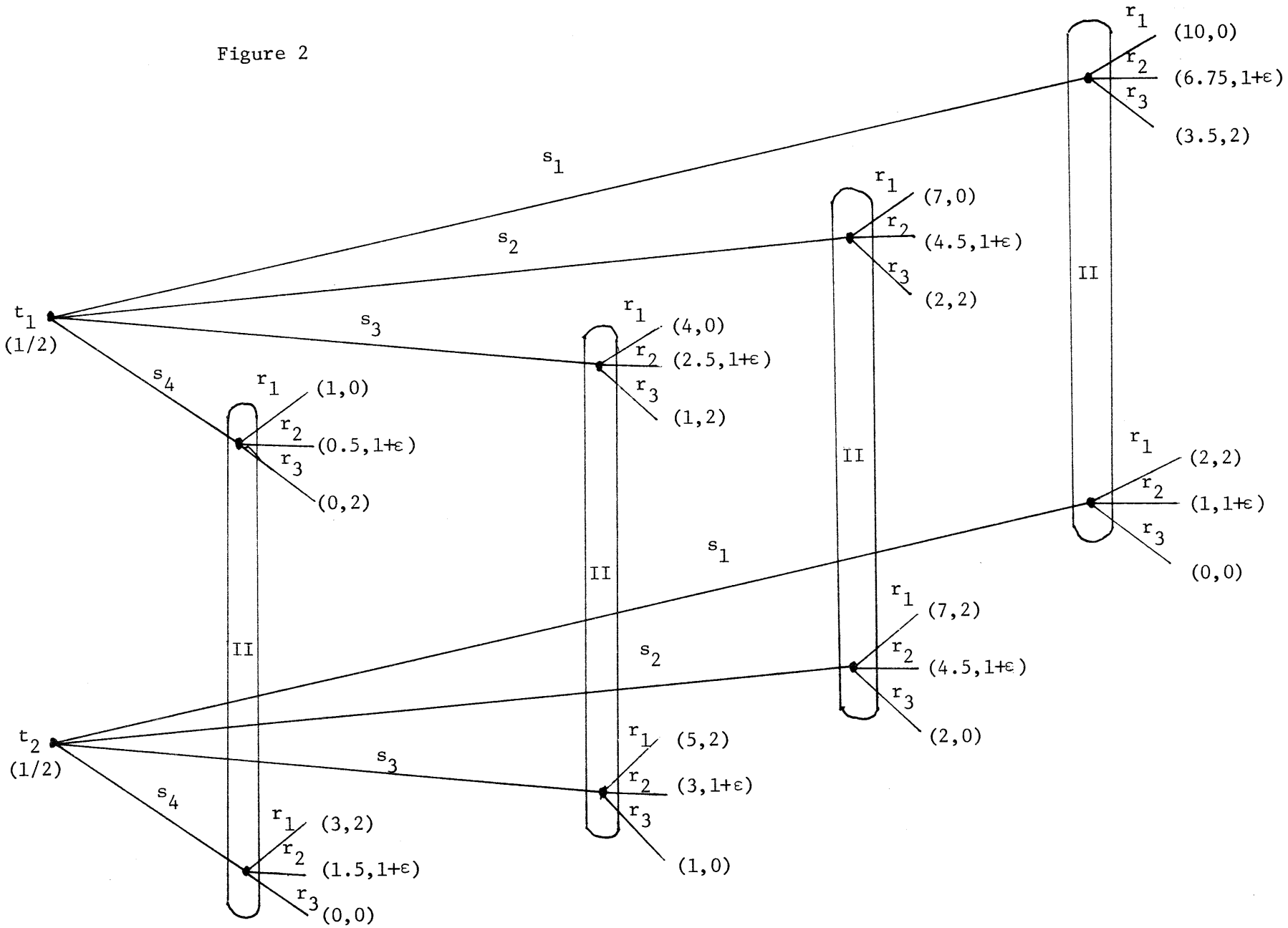

Player II
moves
Player I
moves
1
2
3
4
5
0,3
0,0
0,0
6,2
0,2
t_1
0,3
0,0
0,0
5,3
0,2
0,3
0,0
0,0
0,0
0,2
2,2
0,0
0,3
0,0
0,0
0,2
t_2
0,0
0,3
0,0
6,2
0,2
0,0
0,3
0,0
5,3
0,2
2,2
0,0
0,0
0,3
5,3
0,2
t_3
0,0
0,0
0,3
0,0
0,2
0,0
0,0
0,3
6,2
0,2
2,2

Figure 2


t1
(1/2)
t2
(1/2)
s1
s2
s3
s4
r1
r2
r3
(10,0)
(6.75,1+ε)
(3.5,2)
(7,0)
(4.5,1+ε)
(2,2)
(4,0)
(2.5,1+ε)
(1,2)
(1,0)
(0.5,1+ε)
(0,2)
(2,2)
(1,1+ε)
(0,0)
(7,2)
(4.5,1+ε)
(2,0)
(5,2)
(3,1+ε)
(1,0)
(3,2)
(1.5,1+ε)
(0,0)
II