Economics 703
Advanced Microeconomics **Prof. Peter Cramton**

Lecture Note 4: Refinement[1]

Outline

A. Subgame Perfection Revisited

B. Sequential Equilibrium

   1. Sequential Rationality
   2. Consistency
   3. Structural Consistency

C. Trembling-Hand Perfection

D. Strategic Stability

   1. Backward Induction and Invariance
   2. Admissibility and Iterated Dominance
   3. Stable Equilibrium

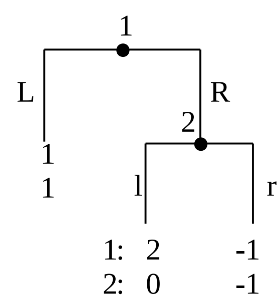
A. Subgame Perfection Revisited

The refinements of the Nash equilibrium concept surveyed in this section are all attempts to extend the spirit of Selten's [1965] definition of subgame perfection to games with imperfect information. To begin, we review the effect of subgame perfection in games of perfect information and the problems that arise with imperfect information.

Game 1: Normal Form

		2: l	2: r
1	L	$1, \underline{1}$	$\underline{1}, \underline{1}$
	R	$\underline{2}, \underline{0}$	-1,-1

Extensive Form



The normal form reveals that both (L,r) and (R,l) are Nash equilibria, but the former relies on an incredible threat by player 2: 2 threatens to play r if player 1 gives 2 the move; this would be bad for 1, so 1 plays L, but it also would be bad for 2. Given the opportunity to move, 2 prefers to play l, so (R,l) is the only sensible Nash equilibrium.

[1] These notes are based without restraint on notes by Robert Gibbons, MIT.

Recall that a *subgame* consists of the game tree following a singleton information set, provided the resulting subtree does not cut any information sets. A Nash equilibrium in the original game is *subgame perfect* if it specifies Nash equilibrium strategies in every subgame.

In game 1, the only proper subgame starts at player 2's decision node. The condition that 2's choice be a Nash equilibrium strategy reduces to a requirement that 2 takes the action that results in the highest payoff. Thus, 2 must play l.

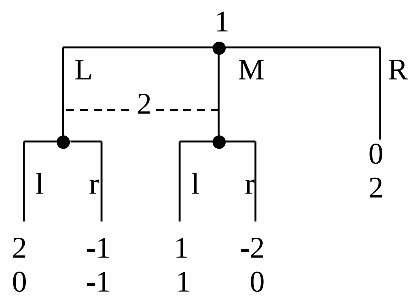
Theorem: Given a finite extensive-form game, there exists a subgame-perfect Nash equilibrium.

The idea of the proof is that a finite game has a finite number of subgames, each of which is finite. Nash's Theorem guarantees that a Nash equilibrium exists for each subgame, and these equilibria can be pieced together, by working backwards through the tree, to construct a subgame-perfect equilibrium. Unfortunately, this process has much less effect in games with imperfect information, as Game 2 illustrates.

Game 2: Normal Form

		2: l	r
	L	$\underline{2}$, $\underline{1}$	-1,-1
1	M	1, $\underline{1}$	-2, 0
	R	0, $\underline{2}$	$\underline{0}$, $\underline{2}$

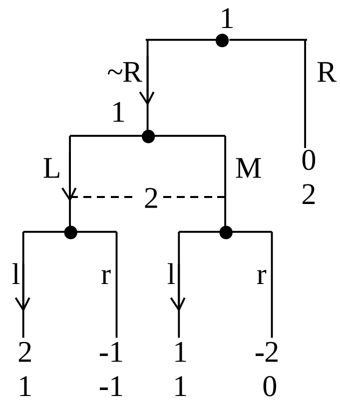
Extensive Form



Here both (L,l) and (R,r) are Nash equilibria, but (R,r) involves an incredible threat by player 2: if 2 gets the move, then r is a strictly dominated strategy for 2, so no matter what player 1 did it is not in 2's interest to play r. And yet, (R,r) is a subgame-perfect equilibrium because there are no (proper) subgames. It would *not* be legitimate to begin a subgame at *one* of player 2's decision notes, thereby cutting his information set, because such a subgame never arises in the original game: at no time does player 2 know which node in the information set has been reached. (This statement is not quite true, since the equilibrium strategies are assumed to be common knowledge. If (L,l) are the equilibrium strategies, then player 2 believes that 1 played L with probability one.)

Another problem with subgame perfection is that the concept is not invariant under inessential transformations of the game tree. For example, the following tree has the same normal form as Game 2, but this time subgame perfection has the desired effect.

Game 2':



In the new game tree, a proper subgame can be defined starting at player 1's second decision node. Since r is a dominated strategy for player 2 it cannot be Nash equilibrium behavior in this subgame, hence (R,r) is not subgame perfect.

What is needed is an extension of the spirit of subgame perfection to games with imperfect information. The definition of a sequential equilibrium involves just such a generalization of backwards induction.

B. Sequential Equilibrium (Kreps and Wilson, *Econometrica* [1982])

1. Sequential Rationality

Kreps and Wilson have proposed "a new criterion for equilibrium in extensive games. The force of this criterion stems from the requirement of *sequential rationality* imposed on the behavior of every player: Every decision must be part of an optimal strategy for the remainder of the game. In games with imperfect or incomplete information, this entails conformity with Savage's axioms of choice under uncertainty: At every juncture the player's subsequent strategy must be optimal with respect to some assessment of the probabilities of all uncertain events, including any preceding but unobserved choices made by other players." (p. 863)

Proceeding informally, consider requiring sequential rationality in Game 2: there must exist beliefs that rationalize each action specified by the strategies. But in game 2 there do not exist beliefs that can rationalize player 2's threat to play r if given the move: l maximizes 2's expected payoff given *any* beliefs about which node has been reached. Therefore, for Game 2, where subgame perfection lost its force, the only Nash equilibrium that is sequentially rational is (L,l).

One perspective on this result is that Kreps and Wilson have simply broadened the definition of a subgame. Instead of starting at a single node, it may now begin with an entire information set. As before, the subgame must contain all the nodes that follow the information set, and no information set can be cut.

Also, a subgame that starts with an information set must specify beliefs about which node in the set has been reached. Sequential rationality imposes the spirit of subgame perfection over this larger class of subgames. In this sense it is a generalized definition of backwards induction. As demonstrated in Game 2, the new formalization of backwards induction prevents the use of strictly dominated strategies off the equilibrium path, whereas Nash equilibria avoid such strategies only on the equilibrium path.

More formally, recall that the information possessed by the players in an extensive-form game is represented in terms of information sets. An information set h for player i is a set of i's decision nodes among which i cannot distinguish. Note that this implies that the same set of actions must be feasible at every node in an information set. Let this set of actions be denoted A(h). Also, let the set of player i's information set be H_i and the set of all information sets be H.

Kreps and Wilson restrict attention to games of perfect recall (a player never forgets what she once knew), for which Kuhn's Theorem implies that mixed strategies can be represented by behavior strategies. A behavior strategy for player i is the collection

$$p_i \equiv \{p^i_h(a)\}_{h \in H_i}$$

where for each $h \in H_i$ and each $a \in A(h)$, $p_h^i(a) \geq 0$ and

$$\sum_{a \in A(h)} p^i_h(a) = 1.$$

That is, $p_h^i(a)$ is a probability distribution that describes i's behavior at information set h. Let the strategy profile be $p = (p^1,\ldots,p^n)$, and let $p^{-i} = (p^1,\ldots,p^{i-1},p^{i+1},\ldots,p^n)$.

We want player i to act optimally at each information set $h \in H_i$, but what is best for i depends on i's beliefs over the nodes in the information set h. Hence, we need to describe player i's beliefs when an information set $h \in H_i$ is reached. Denote these beliefs by $m_h(x)$, where $m_h(x) \geq 0$ is the probability player i assesses that a node $x \in h$ has been reached, and

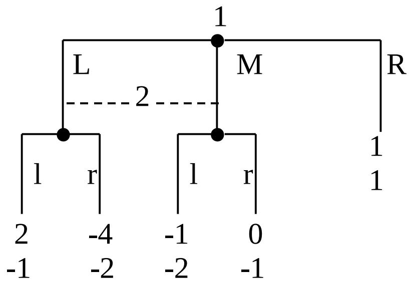
$$\sum_{x \in h} m_h(x) = 1.$$

Let the beliefs throughout the tree be denoted by the collection $m \equiv \{m_h(x)\}_{h \in H}$. Call the beliefs-strategies pair (m,p) an *assessment.*

Definition: An assessment (m,p) is *sequentially rational* if given the beliefs m no player i prefers at *any* information set $h \in H_i$ to change her strategy p_h^i, given the others' strategies p^{-i}.

To summarize, sequential rationality has two effects. First, it eliminates dominated strategies from consideration off the equilibrium path. And second, it elevates beliefs to the importance of strategies. This second effect provides a language—the language of beliefs—for discussing the merits of competing sequentially rational equilibria. As an example of the use of this language, consider Game 3.

Game 3:



There are two sequentially rational Nash equilibria. One is (L,l), which forces the belief at player 2's information set to be that player 1 played L with probability one. The other is (R,r), in which case player 2's information set is not reached in equilibrium. Justifying 2's choice of r requires beliefs about which node in the information set has been reached that make r an optimal decision. A simple calculation shows that if 2 believes that 1 played M with probability $p \geq 1/2$ then r is a best response.

The problem is that this second equilibrium is still implausible: in effect, player 2 is using *implausible beliefs* rather than *incredible actions* to threaten player 1, and the threat succeeds in making 1 play R. The reason 2's beliefs are implausible is that 1 gets a payoff of 1 from playing the equilibrium strategy R. Note that this is superior to any payoff 1 could receive by playing M. And yet when 2's information set is reached 2 believes with probability at least 1/2 that M has been played. Instead, 2 might reason that 1 would not deviate unless there was something to be gained, so L must have been played with probability close to one, in which case l rather than r is the best response.

This kind of argument is sometimes called *forward induction*, or *equilibrium dominance*. (Here R strictly dominated M for player 1, but the same kind of argument applies if the equilibrium payoff to player 1 that follows from R strictly dominates M but other disequilibrium payoffs fail to do so. For instance, there could be a decision node for player 2 following R that pays (1,1) if 2 chooses l and (-1,-1) if 2 chooses r.) See Kreps [1985] and Cho [1986] for more on this subject.

2. Consistency

A *sequential equilibrium* is an assessment (m,p) that is both *sequentially rational* and *consistent*. We now define consistency.

Definition: A strategy profile p is *totally mixed* if it assigns strictly positive probability to each action $a \in A(h)$ for each information set $h \in H$.

Given a totally mixed strategy p, the beliefs m can be derived from Bayes' rule: no part of the tree is reached with prior probability zero, so beliefs conditional on reaching any given information set can be calculated as conditional probabilities.

Definition: An assessment (m,p) is *consistent* if there exists a sequence of totally mixed strategies p_n and corresponding beliefs m_n derived from Bayes' rule such that

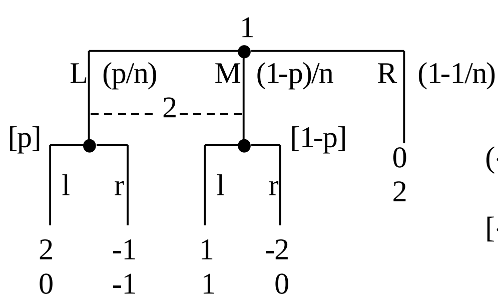
$$\lim_{n\to\infty}(m_n, p_n) = (m,p).$$

Theorem: *For every finite extensive-form game there exists at least one sequential equilibrium. Also, if (m,p) is a sequential equilibrium then p is a subgame-perfect Nash equilibrium.*

It is worth noting that Kreps and Wilson express some doubts about this definition: "We shall proceed here to develop the properties of sequential equilibrium as defined above; however, we do so with some doubts of our own concerning what `ought' to be the definition of a consistent assessment that, with sequential rationality, will give the `proper' definition of a sequential equilibrium." (p. 876)

In many games, the consistency criterion plays no role in the analysis. Consider Game 2, for example. Notice that given the Nash equilibrium strategy vector p = (R,r), *any* belief over the nodes in player 2's information set is consistent. For instance, to verify that the belief (p,1-p) is consistent, consider the totally mixed strategy specified below:

Game 2:



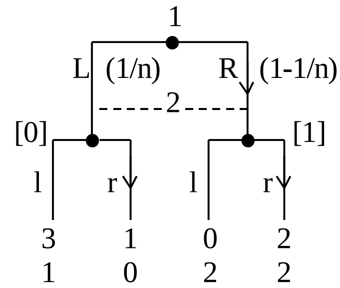
(·) denotes sequence of trembles (p_n),

[·] denotes sequence of beliefs [m_n].

In fact, not only is m the limit of the m_n's derived from the p_n's, but each m_n is (p,1-p). The reason that (R,r) is not a sequential equilibrium strategy, then has nothing to do with a restriction on the beliefs 2 can hold if given the move. Rather, as shown above, it is the failure of r to be sequentially rational for *any* possible belief that causes the problem. This is how sequential equilibrium prevents strictly dominated strategies from being used as threats off the equilibrium path: they are not sequentially rational for any beliefs. Sequential equilibrium does not do nearly as well, however, with weakly dominated strategies.

Game 4:

		2	
		l	r
1	L	$\underline{3}$, $\underline{1}$	1, 0
	R	0, $\underline{2}$	$\underline{2}$, $\underline{2}$



Both (L,l) and (R,r) are Nash equilibria, but r is a weakly dominated strategy for player 2. In other words, the only belief 2 can hold that rationalizes playing r is that player 1 played R with probability one. If 2 assesses any positive probability that 1 played L, then l is the only sequentially rational action. And yet, (R,r) is a sequential equilibrium strategy because the belief {[0],[1]} at 2's information set yields a consistent assessment, as shown in the tree above.

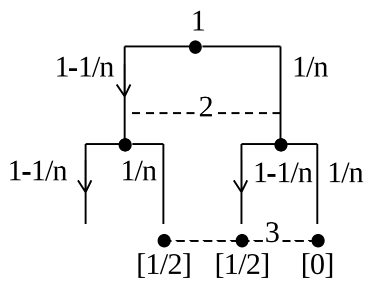
This example contains a general message: sequential equilibrium has (essentially) no cutting power on weakly dominated strategies. Section C shows that Selten's [1975] definition of trembling-hand perfection does eliminate weakly dominated strategies, both on and off the equilibrium path. Unfortunately, this new form of perfection is substantially more complex.

3. Structural Consistency

Since the definition of consistency is purely mathematical, rather than based on economic or strategic thinking, one might ask for a intuitive interpretation of the definition. In this spirit, Kreps and Wilson define *structural consistency*, which they claim is implied by consistency. Unfortunately, their claim is wrong, as demonstrated by Ramey [1985].

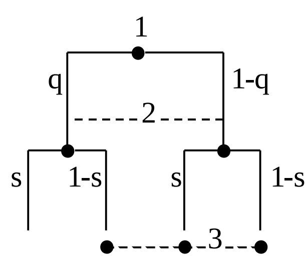
Definition: An assessment (m,p) is *structurally consistent* if for each h $\in$ H there exists a strategy p' that reaches h with positive probability and yields the beliefs $m_h(x)$ over nodes x $\in$ h via Bayes' rule.

If p itself reaches h with positive probability, then $p' = p$ will suffice. The content of the definition applies to information sets that are off the equilibrium path defined by p—information sets p reaches with probability zero. In effect, p' is an alternative hypothesis about how the game has been played that can be adopted when events reject the original hypothesis p. The existence of such an alternative hypothesis seems a better justification for beliefs off the equilibrium path than the continuity property that defines consistency. Unfortunately, the latter does not imply the former, as the following piece of a game tree shows.



Let the equilibrium strategies be indicated by arrows (↓). Then player 3's information set is unreached in equilibrium. Consider the beliefs at this information set given above. Ramey shows that an assessment involving these beliefs and strategies is consistent but not structurally consistent.

To show consistency, let p_n be defined as above. Then m_n at player 3's information set is of the form $(p_n,p_n,1\text{-}2p_n)$ and $p_n \to 1/2$ as $n \to \infty$.



Now try structural consistency. Note that player 2 must make the same decision at each node in her information set since the nodes are indistinguishable (by definition of an information set). Let player 1 play (q,1-q) and player 2 play (s,1-s), as described in the figure above. Then any strategy p' can be represented by the pair $(q,s) \in [0,1]^2$. But the beliefs at 3's information set are then

$$\left[\frac{q(1-s)}{1-qs},\frac{(1-q)s}{1-qs},\frac{(1-q)(1-s)}{1-qs}\right],$$

and the only way to get zero probability on the third node is q=1 or s=1. We cannot have q=s=1 because then the information set is not reached with positive probability, as required by structural consistency. But if only one of q and s is one then the probability of reaching either the first or the second node is zero, not 1/2 as required.

The most important implication of Ramey's example is that the appeal of the definition of consistency is reduced. Recall that Kreps and Wilson explicitly expressed doubts about it in the paper. Ramey's example provides a fresh and important reason for doubt.

One might think of dropping consistency and requiring only structural consistency. Unfortunately, Kreps and Wilson point out (on p. 876) that such an approach does not even guarantee *subgame* perfection.

C. Trembling-Hand Perfection (Selten [1975])

The behavioral strategy p is a *trembling-hand-perfect equilibrium* if there exists a sequence of totally mixed strategies p_n such that

(1) $p_n \rightarrow p$, and

(2) for each i *and each n*, p^i is a best response to p_n^{-i}.

Theorem: *For every finite extensive-form game there exists at least one trembling-hand-perfect equilibrium. Also, if* p *is a trembling-hand-perfect equilibrium then there exists beliefs* m *such that* (m,p) *is a sequential equilibrium.*

To compare trembling-hand-perfect and sequential equilibria, notice that the totally mixed strategies p_n implicitly generate beliefs m_n, via Bayes' rule. Let m be the limit of the sequence m_n. In terms of assessments, (m,p) is trembling-hand perfect if there exists a sequence of totally mixed strategies p_n and associated beliefs m_n such that

(1a) $(m_n,p_n) \rightarrow (m,p)$,

(1b) (m,p) is sequentially rational, and

(2) for each i and n, p^i is a best response to p_n^{-i}.

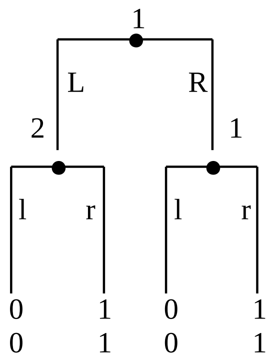
Notice that (1a) and (1b) are consistency and sequential rationality, the defining properties of a sequential equilibrium. This proves the second sentence of the Theorem. The important difference is (2): sequential rationality is required only *at* the limit for a sequential equilibrium, but *also on the way to* the limit for a trembling-hand-perfect equilibrium.

The costs and the benefits of this definition of perfection both follow from condition (2). On the one hand, it eliminates weakly dominated strategies. But on the other, it is vastly harder to verify than (1a) and (1b). To appreciate the effect of (2), return to Game 4. Recall that r is only a best response for player 2 if she believes that player 1 has played R with probability one; if 2 assesses any positive probability that 1 played L then l, not r, is the best response. Thus, r fails (2).

One might suspect that because of the continuity built into the consistency condition (1a), sequential rationality at the limit might imply sequential rationality sufficiently near the limit. In fact, this is almost always true. More precisely, in terms of Lebesque measure on the space of payoffs, the two concepts are almost surely identical: for almost every game, every sequential equilibrium is trembling-hand perfect.

This result may seem more reassuring than it is: weakly dominated strategies have Lebesque measure zero in payoff space, but may well have positive "economic" measure, and it is precisely on the question of weak dominance that the concepts differ.

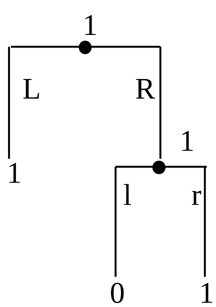
Perfection can be defined for normal form games as well. For a particular extensive-form game, how does the set of normal-form perfect equilibria compare with the set of extensive-form perfect equilibria. One might suspect that every extensive-form perfect equilibrium is normal-form perfect since the extensive form contains more information and so may have more sequential rationality constraints. This turns out to be false as the following example illustrates:



	l	r
Ll	0	1
Lr	0	1
Rl	0	0
Rr	1	1

Here (Rr,r) is the unique normal-form perfect equilibrium, but (Lr,r) is an extensive-form perfect equilibrium. Rr weakly dominates all of 1's other strategies in the normal form, but in the extensive form Lr does just as well Rr, since 1 is just as worried about her own tremble after moving R as 2's tremble after moving L.

Is every normal-form perfect equilibrium an extensive form perfect equilibrium? This is certainly not the case as shown in the following example:



L	1
Rl	0
Rr	1

Here L is the unique extensive-form perfect equilibrium, but Rr is a normal-form perfect equilibrium as well. Rr is worse than L in the extensive form, because 1 has to be concerned about a tremble after moving R. But L and Rr are equivalent pure strategies in the normal form.

D. Strategic Stability (Kohlberg and Mertens, *Econometrica* [1986])

Which Nash equilibria are strategically stable; that is, once a particular strategy profile is specified, is it the case that no player can benefit from unilaterally deviating from his prescribed strategy? Does every game have a strategically stable equilibrium? Kohlberg and Mertens propose a set-valued equilibrium concept in an effort to shed some light on these questions. They take the traditional point of view that the strategic possibilities of the players and the consequences of any feasible behavior are completely represented by the noncooperative game (either in extensive or normal form). There is no communication during the play of the game, but the game is preceded by communication to identify a particular equilibrium. The paper ignores the process by which a candidate equilibrium is arrived at, focusing solely on the question of once a profile of strategies is isolated is it the case that the strategies are self-enforcing (no one wants to unilaterally deviate).

We begin with some definitions.

The *agent normal form* of a game tree is the normal form of the game between *agents*, obtained by letting each information set be manned by a different agent where we give any agent of the same player that players' payoff.

A *behavioral strategy* describes what to do at each information set (i.e., a (mixed) strategy for each of a player's agents).

A *sequential equilibrium* of an extensive form game (Kreps and Wilson) is an n-tuple of behavior strategies which is a limit of the sequence (s_m) of completely mixed behavior strategies such that every agent maximizes his payoff given strategy and beliefs at each information set implied by the limit of (s_m).

An *e-perfect equilibrium* of a normal form game (Selten) is a completely mixed strategy vector, such that any pure strategy which is not a best reply has weight less than ε.

An *e-proper equilibrium* of a normal form game is a completely mixed strategy vector, such that whenever some pure strategy s_1 is a worse reply than some other pure strategy s_2, the weight on s_1 is smaller than ε times the weight on s_2.

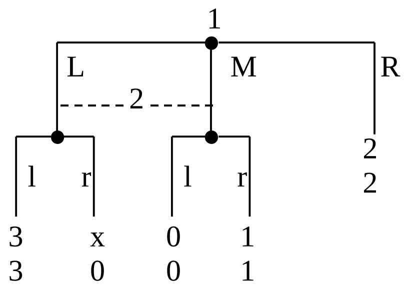
A *perfect (proper) equilibrium of a normal form game* is a limit ($e \to 0$) of e-perfect (proper) equilibria.

A *perfect (proper) equilibrium of a tree* is a perfect (proper) equilibrium of its agent normal form.

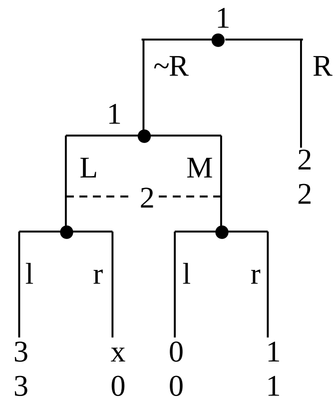
1. Backwards Induction Rationality and Invariance

A necessary condition for stability is sequential (backwards induction) rationality: the players act optimally at every point in the game tree. Sequential, perfect, and proper equilibria satisfy this condition. One of the problems of sequential equilibria is that the set of sequential equilibria is not invariant to inessential transformations of the game tree as illustrated below, where $1 < x \leq 2$.

Game 5:



Game 5':



Although Game 5 and Game 5' are essentially the same (they have the same normal form), (R,r) is a sequential equilibrium in Game 5, but is not a sequential equilibrium in Game 5'. It seems natural to require that a strategically stable equilibrium of a game tree must be sequential in any other game tree having the same normal form. Is such a requirement too strong to guarantee existence? The following proposition confirms that the answer is no (since a proper equilibria always exists).

Proposition 0. A proper equilibrium of a normal form is sequential in any tree with that normal form.

Proof: Let x = lim x_e be a proper equilibrium where the x_e are e-proper equilibria. Given a tree, let s_e be behavioral strategies equivalent to x_e and let m_e be the vector of conditional probabilities they imply on information sets, and let $s_e \to s$ and $m_e \to m$. We have to show that s is such that each agent maximizes her payoff given m and the strategies of the others.

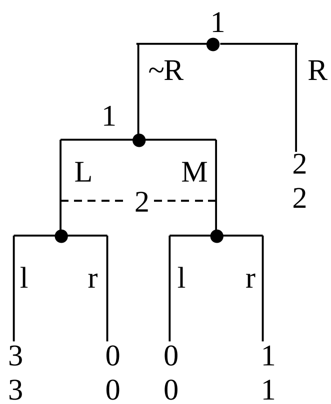
Assume the contrary. Then there is some player, say 1, and a last information set for her, say J, such that s^1 assigns positive probability to a move in J, say L, whose expected payoff (given m and s) is less than that of another move, say R. Clearly, the same is true for m_e and s_e, provided $e > 0$ is sufficiently small.

It follows that every normal form strategy of 1 that does not avoid J and chooses L in J has a smaller expected payoff, given x_e, than a modification of that strategy that chooses R and then continues as in s^1. Since x_e is e-proper, x_e^1 assigns the first strategy less than e times the probability of the second strategy. It follows that s_e^1 assigns to L a probability of at most ke, where k is the number of normal-form strategies of 1. Letting $e \to 0$, we see that s^1 assigns to L zero probability, a contradiction.¦

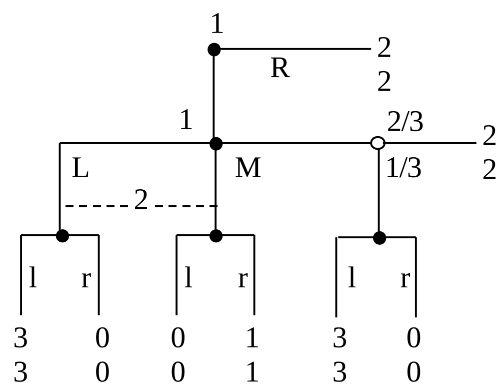
This is an important result because it goes against the notion that only the extensive form contains the detail necessary to determine if an equilibrium is sequential. By restricting attention to proper equilibria of the normal form we can guarantee sequential rationality in the (any) extensive form. Although this result casts doubt on the claim that it is *essential* to analyze the extensive form, rather than the normal form, there is still good reason to be be interested in the extensive form. Namely, for non-trivial dynamic games the set of behavioral strategies in the extensive form is of much lower dimension than the set of mixed strategies in the normal form.

A further invariance requirement is that strategically stable equilibria depend only on the *reduced normal form*, determined by eliminating all pure strategies that are convex combinations of other pure strategies.

Game 6:



Game 6':



Game 6' is the same as Game 6 with the explicit inclusion of the mixed strategy in which R is played with probability 2/3 and L is played with probability 1/3. Notice that both (L,l) and (R,r) are sequential equilibria in Game 6, but in 6' (L,l) is the only sequential equilibrium: the randomized strategy strictly dominates M for player 1, so that 2 must but zero weight on node M and so 2 must play l in which case 1 plays L.

2. Admissibility and Iterated Dominance

Strategies that are not *(weakly) dominated* are called *admissible*. Admissibility is a natural requirement of stability. Why should a player employ a dominated strategy? Both perfect and proper equilibria satisfy admissibility (dominated strategies are never played with positive probability on or off the equilibrium path), but sequential equilibria can involve weakly dominated strategies.

But if we are eliminating dominated strategies, shouldn't we be eliminating dominated strategies iteratively and require that the deletion of such strategies can have no impact on strategic stability? The problem with requiring *iterated dominance* of this form is that it is incompatible with existence, as seen in the following example.

	L	R
U	$\underline{3}$, $\underline{2}$	$\underline{2}$, $\underline{2}$
M	1, $\underline{1}$	0, 0
D	0, 0	1, 1

The equilibrium (either (3,2) or (2,2)) found by iterated dominance depends on the order in which dominated strategies are eliminated. Both (2,2) and (3,2) are inadmissible. In order to get existence, it is necessary to *include* all equilibria that satisfy iterative dominance, rather than *exclude* equilibria that can be eliminated by iterative dominance.

3. Stable Equilibria

The desireable properties for a set-valued definition of stability can be summarized as follows:

Existence: Every game has at least one solution.

Connectedness: Every solution is connected; i.e., the solution is a connected set in the simplex of mixed strategies.

Backwards Induction: A solution of a tree contains a backwards induction (either sequential or perfect) equilibrium of the tree.

Invariance: A solution of a game is also a solution of any equivalent game; i.e., having the same reduced normal form.

Admissibility: The players' strategies are undominated at any point in a solution.

Iterated Dominance: A solution of a game G contains a solution of any game G' obtained from G by deletion of a dominated strategy.

Forward Induction: A solution must remain so after deletion of a strategy which is an inferior response for all strategies contained in the solution.

A game is perturbed by either changing the payoffs a small amount or changing the strategies a small amount. The definitions of stability involve continuity properties of the set of Nash equilibria for perturbed games.

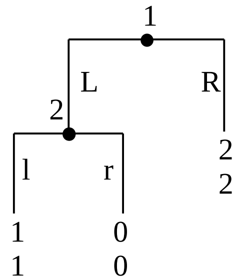
Proposition 1. The set of Nash equilibria of any game has finitely many connected components. At least one is such that for any equivalent game (i.e., same reduced normal form) and for any perturbation of the normal form there is a Nash equilibrium close to this component.

Hyperstable Equilibrium:

S is a *hyperstable set of equilibria* in a game G if it is minimal with respect to the following property:

Property (H): S is a closed set of Nash equilibria of G such that, for any equivalent game, and for any perturbation of the normal form of that game, there is a Nash equilibrium close to S.

Hyperstable equilibria allow any perturbation of the game (both payoffs and strategies). A hyperstable equilibrium satisfies existence, a version of connectedness, backwards induction, invariance, and iterated dominance, but does not satisfy admissibility, as seen by the following example:



	l	r
R	$\underline{2}$, $\underline{2}$	$\underline{2}$, $\underline{2}$
L	1, $\underline{1}$	0, 0

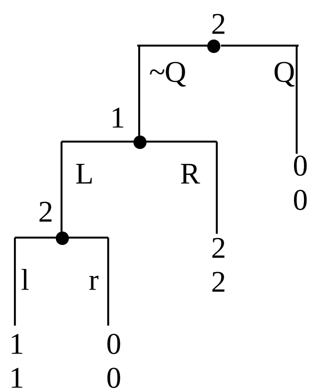
The unique hyperstable set is the full interval from (R,l) to (R,r), but only (R,l) is admissible. Any mixture of l and r for player 2 must be included in the hyperstable set, since we can perturb the game by adding that mixture as a pure strategy and then increase its payoff slightly. (This example also illustrates how a generic extensive form can easily lead to a non-generic normal form.)

Fully Stable Equilibrium:

S is a *fully stable* set of equilibria of a game G if it is minimal with respect to the following property:

Property (F): S is a closed set of Nash equilibria of G satisfying: for any $\varepsilon>0$ there exists a $\delta>0$ such that, whenever each player's strategy set is restricted to some compact convex polyhedron contained in the interior of the simplex and at (Hausdorff) distance less than δ from the simplex, then the resulting game has an equilibrium point ε-close to S.

A fully stable equilibrium allows any perturbation of the players strategies, but the payoffs cannot be perturbed. A fully stable equilibrium satisfies existence, connectedness, backwards induction, invariance, and iterated dominance, but fails to satisfy admissibility:



	l	r	Q
R	$\underline{2}$, $\underline{2}$	$\underline{2}$, $\underline{2}$	$\underline{0}$, 0
L	1, $\underline{1}$	0, 0	$\underline{0}$, 0

Here only (R,l) is admissible, but the unique fully stable set is the interval from (R,l) to (R,r). For example, (R,r) must be included, since it is the unique equilibrium when l is perturbed more than r toward Q.

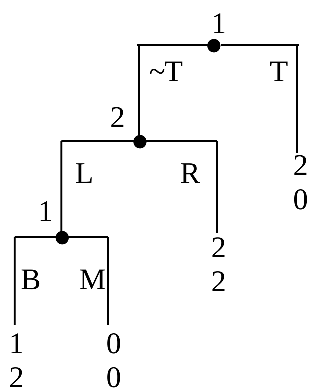
Stable Equilibria:

A set S of equilibria is *stable* in a game G if it minimal with respect to the following property:

Property (S): S is a closed set of Nash equilibria of G satisfying: for any $e>0$ there exists $\delta_0>0$ such that for any completely mixed strategy vector $s_1,..., s_n$ and for any $\delta_1,...,\delta_n$ with $0 < \delta_i < \delta_0$, the perturbed game where every strategy s of player i is replaced by $(1-\delta_i)s + \delta_i s_i$ has an equilibrium e-close to S.

This is the same definition as full stability, but perturbed strategies are restricted to simplices with faces parallel to the faces of the original simplex. If we say "some" instead of "any" $s_1,..., s_n$ and $\delta_1,...,\delta_n$, then we get perfect equilibria. A stable equilibrium satisfies existence, a version of connectedness, invariance, admissibility, iterated dominance, and forward induction. Unfortunately, a stable equilibrium need not satisfy connectedness or backward induction.

The following example illustrates the tension between forward induction and backward induction and the importance of a set-valued equilibrium concept to resolve this tension.



	L	R
T	$\underline{2}$, $\underline{0}$	2, $\underline{0}$
M	0, 0	$\underline{3}$, $\underline{1}$
B	1, $\underline{2}$	$\underline{3}$, 1

(T,L) is the unique backward induction equilibrium, but if we delete row B since it is an inferior response at this equilibrium, then (M,R) becomes the forward induction equilibrium. The unique stable set is (T,L) to (T, 1/2 L). The equilibrium outcome is always (2,0), but row B is no longer inferior to all points in the stable set, so forward induction is satisfied.