CUR 412: Game Theory and its Applications, Lecture 13

Prof. Ronaldo CARPIO

May 31, 2016

Announcements

- Homework #4 is due today.
- Homework #5 will be posted on the web site later today, due in two weeks.

Review of Last Lecture

- A *signaling game* is a game where one player has private information that is not observed by other players.
- This player can choose an action that may depend on his private information; other players observe this action.
- In a *separating equilibrium*, the first player's action reveals his private information; in a *pooling equilibrium*, it does not.
- For example, suppose the private information is the *type* of Player 1, chosen by Nature according to some probability distribution.
- In a separating equilibrium, different types of Player 1 choose different actions, so they can be distinguished.
- In a pooling equilibrium, different types of Player 1 choose the same action.

Ch. 10.7: Education as a signal of ability

- Why do students obtain a college degree?
- One reason is that the knowledge they gain in college will increase their skills and abilities.
- However, there is another possible reason: perhaps students use degrees to *differentiate* themselves from other students when applying for jobs.
- This can hold even if the degree itself does not increase ability.
- We model this as a signaling game.

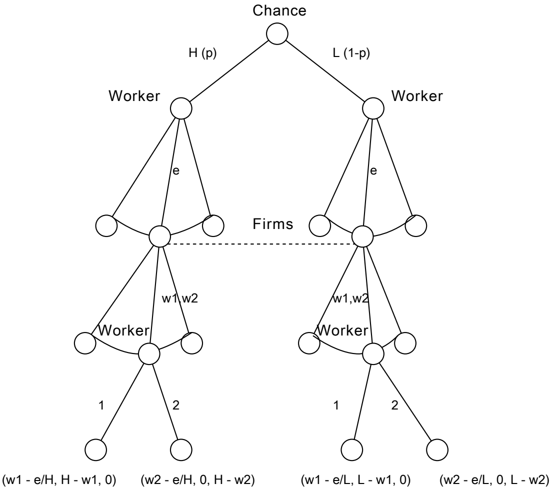
Ch. 10.7: Education as a signal of ability

- Suppose the ability level of a worker can be measured by a single number
- There are two types of workers: "high" and "low"-ability workers, denoted H and L, with $L < H$.
- Type is known to the workers, but cannot be directly observed by employers.
- Workers can choose to obtain some amount of education, which has no effect on ability, but costs less for the H-type worker.

Ch. 10.7: Education as a signal of ability

- The sequence of the game is as follows:
  - Chance chooses the type of the worker at random; the probability of H is p.
  - The worker, who knows his type, chooses an amount of education $e \geq 0$. The cost of education is different according to type; for a L-type worker, the cost is e/L; for a H-type worker, it is e/H.
  - Two firms observe the worker's choice of e (but not his type), and simultaneously offer two wages, w_1 and w_2.
  - The worker chooses one of the wage offers and works for that firm. The worker's payoff is his wage minus the cost of education. The firm that hires the worker gets a payoff of the worker's ability, minus the wage. The other firm gets a payoff of 0.

Game Tree


Chance
H (p)
L (1-p)
Worker
Worker
e
e
Firms
w1,w2
w1,w2
Worker
Worker
1
2
1
2
(w1 - e/H, H - w1, 0)
(w2 - e/H, 0, H - w2)
(w1 - e/L, L - w1, 0)
(w2 - e/L, 0, L - w2)

Finding WSE

- We claim there is a weak sequential equilibrium in which a H-type worker chooses a positive amount of education, and a L-type worker chooses zero education.
- Consider this assessment (i.e. beliefs plus strategies), where e^* is a positive number (to be determined):
  - *Worker's strategy*: Type H chooses $e = e^*$ and type L chooses $e = 0$. After observing $w1, w2$, both types choose the highest offer if $w1, w2$ are different, and firm 1 if they are the same.
  - *Firms' belief*: Each firm believes that a worker is type H if he chooses $e = e^*$, and type L otherwise.
  - *Firms' strategies*: Each firm offers the wage H to a worker who chooses $e = e^*$, and L to a worker who chooses any other value of e.

Finding WSE

- Let's check that the conditions for consistency of beliefs and optimality of strategies are satisfied.
  - Consistency of beliefs: take the worker's strategy as given.
  - The only information sets of the firm that are reached with positive probability are after $e = 0$ and $e = e^*$; at all the rest, the firms' beliefs may be anything.
  - At the information set after $e = 0$, the only correct belief is $P(H|e = 0) = 0$.
  - At the information set after $e = e^*$, the only correct belief is $P(H|e = e^*) = 1$. So these beliefs are consistent.
  - Optimality of firm's strategy: Each firm's payoff is 0, given its beliefs and strategy.
  - If a firm deviates by offering a higher wage, it will make a negative profit.
  - If it deviates by offering a lower wage, it gets a payoff of 0 since the worker will choose the other firm. So, there is no incentive to deviate.

Finding WSE

- Optimality of worker's strategy: In the last subgame, the worker's strategy of choosing the higher wage is clearly optimal. Let's consider the worker's choice of e:
  - Type H: If the worker maintains the strategy and chooses $e = e^*$, he will get a wage offer of H and his payoff will be $H - \frac{e^*}{H}$.
  - If the worker deviates and chooses any other e, he will get a wage offer of L and his payoff will be $L - \frac{e}{H}$.
  - The highest possible payoff when deviating is when $e = 0$, which gives a payoff of L.
  - Therefore, in order for our hypothetical equilibrium to be optimal, we need $H - \frac{e^*}{H} \geq L$, or

$$e^* \leq H(H - L)$$

Finding WSE

- Type L: If the worker maintains the strategy and chooses $e = 0$, he will get a wage offer of L and his payoff will be L.
- If the worker deviates and chooses anything but e^*, he still gets a wage offer of L and a lower payoff of $L - \frac{e}{L}$.
- If the worker deviates and chooses e^* (i.e. imitates a H-type) then he gets a wage offer of H, for a total payoff of $H - \frac{e^*}{L}$.
- For our hypothetical equilibrium to be optimal, we need $L \geq H - \frac{e^*}{L}$ or

$$e^* \geq L(H - L)$$

Conditions for Equilibria

- Combining these requirements, the condition for this equilibrium to be optimal is:

$$L(H - L) \leq e^* \leq H(H - L)$$

- If this is satisfied, then separating equilibria exist in which H-type workers can be distinguished from L-type workers by their choice of e.
- This is not the only type of equilibrium that exists: there may also exist pooling equilibria, given the same values of H and L, in which both types of workers choose the same amount of education.

- Note that in order for a separating equilibrium to exist, the signal must be *costly* to the sender. Otherwise, the L-type can always imitate the H-type.
- A signal that is not costly is called *cheap talk*.
- Applications of signaling games in biology: the *handicap model*
- In some animal species, the male develops seemingly useless and costly features.
- For example, the antlers of stags, the tail of peacocks, etc.
- Biologists have developed models where these are signals of genetic fitness.

Chapter 14: Repeated Prisoner's Dilemma

	C	D
C	2,2	0,3
D	3,0	1,1

- Let's recall the Prisoner's Dilemma.
- Here, we are labeling the actions for each player as *Cooperate* or *Defect*.
- As we have seen, this game has a single Nash equilibrium (D, D), where both players choose *Defect*.
- In real life, however, people frequently manage to sustain cooperation, in contrast to the theoretical prediction of the Prisoner's Dilemma model.

- One possible explanation for this is that playing Prisoner's Dilemma only once misses a key feature of the real world: that agents interact *repeatedly*.
- If agents know they will interact again in the future, defecting in one time period may be punished by reciprocal defecting in the future.
- Agents can develop a *reputation* for cooperating or defecting.
- We will study a specific case of repeated interaction, where the same agents meet in several periods and play the Prisoner's Dilemma in each period.
- In order to analyze this situation, we will need to define what a strategy is, and what preferences are for games played in several periods.

Discounting Future Payoffs

- How should a player evaluate a sequence of payoffs in different time periods?
- Suppose that Player i is playing a repeated game for T periods, and that his payoff in each period is given by the sequence of values

$$w^1, w^2, w^3, ..., w^T$$

- w^j is the payoff in the jth period.
- We will assume a type of preference called *discounting*, where *payoffs in the future* are **less valued** than *payoffs in the present* by a constant factor.
- To be specific, we will assume that every player has a *discount factor* δ, which is between 0 and 1.

Discounting Future Payoffs

- Let's compare two streams of payoffs: $w^1, w^2, ...w^T$ and $v^1, v^2, ...v^S$ where T and S may be different, or can be an infinitely long sequence.
- We will assume that the player prefers the stream that has the highest *discounted sum*:

$$w^1 + \delta w^2 + \delta^2 w_3 + ...\delta^{T-1} w^T = \sum_{k=1}^{T} \delta^{k-1} w^k$$

$$v^1 + \delta v^2 + \delta^2 v_3 + ...\delta^{S-1} v^S = \sum_{k=1}^{S} \delta^{k-1} v^k$$

Discounting Future Payoffs

- A player with a discount factor between 0 and 1 is said to be *impatient*
- A lower value of δ implies the player is *more* impatient, since the player puts less weight on payoffs in the future, relative to payoffs in the present.
- If the discount factor is 0, then the player is completely *myopic* (i.e. short-sighted), and does not care about the future at all.
- If the discount factor is 1, then the player places equal value on payoffs today and payoffs in the future.

Geometric Series

- It will be useful to know the formulas for the sum of a geometric series.
- Suppose $0 < r < 1$. Let S denote the infinite sum

$$S = 1 + r + r^2 + \ldots = \sum_{t=0}^{\infty} r^t$$

- Then $rS = S - 1$, therefore $S = \frac{1}{1-r}$.
- We can also find the sum of a finite series:

$$1 + r + \ldots + r^T = \sum_{t=0}^{T} r^t$$

$$= \sum_{t=0}^{\infty} r^t - \sum_{t=T+1}^{\infty} r^t = S - r^{T+1} S = S(1 - r^{T+1})$$

Discounting Future Payoffs

- Why should people value payoffs in the future less than payoffs in the present?
- One answer is that in the real world, there is always a positive chance of death in any given time period.
- Suppose a person will retire in T years, at which point he will get a payoff of w.
- In each year, the person will survive to the next year with probability p, or exits the game with probability $1 - p$.
- If the person exits before the T-th period, he gets a payoff of 0.

Discounting Future Payoffs

- The probability of surviving the first period is p; the probability of surviving 2 periods is p^2, and so on.
- The probability of surviving T periods is p^T, so the expected payoff at the beginning of the game is $p^T w$.
- A discount rate of δ corresponds to a probability of exit of $1 - \delta$ in every period.

Discounting Future Payoffs

- Another explanation for discounting is that the market offers a positive interest rate, and any payoffs need to be compared to what could be earned at the market interest rate.
- Suppose the market interest rate is r, so if 100 is invested in the bank, it will return $(100)(1+r)$ one period in the future, $(100)(1+r)^2$ two periods in the future, etc.
- Let's compare the *net present value* of a payoff of 100 in the present, vs. 100 one period in the future.
- If $\frac{100}{1+r}$ is invested in the bank today, it will return 100 one period in the future; so this is equivalent to a discount factor of $\frac{1}{1+r}$.

Discounting Future Payoffs

- Suppose you had a security that paid 100 every year for perpetuity. Its present value would be:

$$100 + \frac{100}{1+r} + \frac{100}{(1+r)^2} + \ldots = 100\sum_{t=0}^{\infty}\frac{1}{(1+r)^t}$$

$$= \frac{100}{1 - \frac{1}{1+r}} = 100\frac{1+r}{r}$$

- A discount rate of δ corresponds to an interest rate $1 + r = 1/\delta \rightarrow r = 1/\delta - 1$.

Discounted Average

- Finally, for convenience, we would like a player to be indifferent between a one-time payoff of c in the present, and the infinite stream of constant payoffs $(c, c, c, \dots)$.
- The discounted sum of the infinite sequence $(c, c, c, \dots)$ is:

$$c + \delta c + \delta^2 c + \dots = c(1 + \delta + \delta^2 + \dots) = c\frac{1}{1-\delta}$$

- The discounted sum of a *finite* constant sequence is:

$$c + \delta c + \delta^2 c + \dots + \delta^k c = c\sum_{i=0}^{k} \delta^i$$

$$= c\left[\sum_{i=0}^{\infty} \delta^i - \delta^{k+1}\sum_{i=0}^{\infty} \delta^i\right] = c\left[\frac{1}{1-\delta} - \frac{\delta^{k+1}}{1-\delta}\right]$$

- We can make the player indifferent between c and $(c, c, \dots)$ by multiplying by $1 - \delta$. This is called the **discounted average**:

$$(1-\delta)\sum_{t=1}^{\infty} \delta^{t-1} w^t$$

Discounted Average

$$(1-\delta)\sum_{t=1}^{\infty}\delta^{t-1}w^{t}$$

- For the rest of the repeated games topic, we will assume preferences are represented by the discounted average of a sequence of payoffs.

Repeated Games

- Suppose G is a strategic game. We define the T-period repeated game of G as an extensive game with perfect information and simultaneous moves in which:
  - Players: the set of players is the same as in G
  - Terminal histories: the set of terminal histories is the set of all possible sequences $(a^1, a^2, ...a^T)$, where a^k is an action profile in G. If there are N possible action profiles of G, then there are N^T possible terminal histories.
  - Player function: all players move after every history.
  - Actions: the set of actions after every history is A_i, the same action set in G
  - Preferences: Each player evaluates the terminal history $(a^1, a^2, ...a^T)$ by the discounted average of the payoffs resulting from the outcomes in the terminal history:

$$(1-\delta)\sum_{t=1}^{T}\delta^{t-1}u_i(a^t)$$

Repeated Games

- In short, the T-period repeated game of G is the situation where in each period, both players play G; payoffs are the discounted average of the length-T sequence of payoffs from each period.
- The *infinitely* repeated game of G is the same except that terminal histories are now infinite sequences, and preferences are represented by the discounted average of an infinite series of payoffs.

Strategies in a Repeated Game

- As we've seen before, a strategy in an extensive game needs to specifies an action after every history in which it is the player's turn to move.
- In repeated games, all players move after every history, so a strategy must specify a player's action after any possible history.
- In an infinitely repeated game, this could potentially require specifying actions for all possible histories of any length.
- We can simplify things by only looking at a special class of strategies, in which actions can depend only a finite subset of the past history.
- In other words, the strategy has a limited "memory", and can only "remember" a finite number of past moves.
- Here are some examples of this kind of strategy:

Grim Trigger

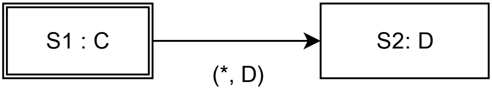
- **Grim Trigger**: This is a simple strategy that always plays C until the other player plays D; then it "punishes" the other player by always playing D.
- Formally, we define this strategy as:

$$s_i(\emptyset) = C$$

$$s_i(a^1, \ldots, a^t) = \begin{cases} C & \text{if } (a^1, \ldots, a^t) = (C, C, \ldots, C) \\ D & \text{otherwise} \end{cases}$$

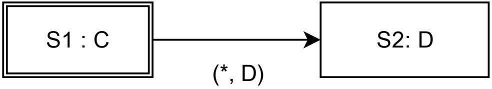
- The first part of the definition, $s_i(\emptyset) = C$, specifies what to do at the beginning of the game.
- The second part specifies what to do for any finite history.
- If the other player has never played D, then play C; if the other player has played D at any point in time, then play D.

Grim Trigger



- We can graphically represent strategies using a *state diagram*.
- Each box in this diagram represents a possible *state* of the strategy; here, there are two states: $S1, S2$.
- At the beginning of the game, the strategy starts out in the box with double edges, $S1$. The last word in the box, C, specifies what action to play in this state.

Grim Trigger



- Then, depending on what the outcome of the game in this period is, the strategy will either move to another state, or remain in the same state.
- The arrow labeled $(*, D)$ specifies when to *transition* to another state: if the outcome of the game is $(*, D)$, move to state $S2$.
- The $*$ means that any action of the first player, together with D played by the second player, will trigger this transition. If the outcome does *not* match $(*, D)$, then the strategy will remain in state $S1$.
- Once in state $S2$, the specified action is always D, and there are no more transition arrows out of this state, which means that the strategy will remain in this state forever (therefore, play D forever).

Always Cooperate

- **Always Cooperate**: This is one of the simplest strategies; it plays C after any history. Formally, it is defined as:

$$s_i(\varnothing) = C$$

$$s_i(a^1, \ldots, a^t) = C$$

- In the state diagram, the strategy begins in state $S1$, and remains there, always playing C.

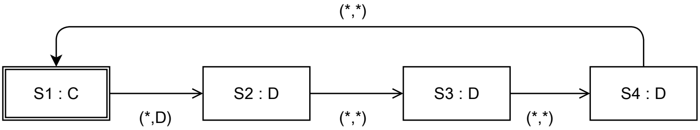
Always Defect

- **Always Defect**: Likewise, this strategy is defined as:

$$s_i(\varnothing) = D$$

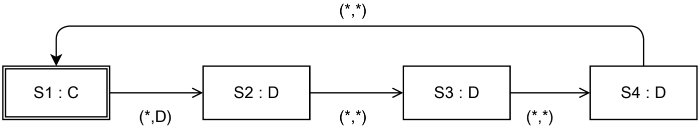
$$s_i(a^1, ..., a^t) = D$$

Punish for 3 periods



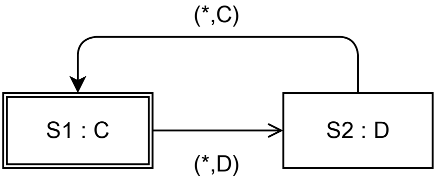
- Here's a an example of a more complicated strategy.
- **Punish for 3 periods**: This strategy plays C until the other player plays D, at which point this strategy will play D for 3 consecutive periods. Then, this strategy will "forget" the past and go back to its original state.
- This is complicated to define as a function, but relatively simple as a diagram.
- This strategy begins at state $S1$, which plays C.

Punish for 3 periods



- If the other player plays D:
  - then this strategy transitions to state $S2$, which plays D once;
  - then to state $S3$, which plays D once;
  - then to state $S4$, which plays D once;
  - then transitions back to the original state $S1$.
- Thus, this strategy will punish D by playing D 3 times, after which it will play C again (even if the other player previously played D in the last period of punishment).

Tit-for-Tat



- **Tit-for-Tat**: This strategy has an intuitive interpretation: do whatever the other player did previously. We can define it as:

$$s_i(\varnothing) = C$$

$$s_i(a^1, \ldots, a^t) = \begin{cases} C & \text{if } a^t = C \\ D & \text{if } a^t = D \end{cases}$$

Finitely repeated Prisoner's Dilemma

- In a one-shot Prisoner's Dilemma, the NE is when both players *Defect*.
- Can a *finitely repeated* Prisoner's Dilemma sustain a different NE?
- Suppose Player 1's strategy is s_1 and Player 2's strategy is s_2.
- Let t denote the last period in which the outcome is *not* (D, D) (and therefore the outcome in all periods after t is (D, D)).
- Suppose that Player 1 chose C in this period (we could also assume it was Player 2).
- We claim that Player 1 can deviate and get a higher payoff.

Finitely repeated Prisoner's Dilemma

- Let s_1' be any strategy such that the strategy profile (s_1', s_2) results in exactly the same history as (s_1, s_2), except that Player 1 chooses D in period t.
- This must increase Player 1's payoff in period t, while his payoff in periods *after* t cannot be worse, since Player 1 is already playing D in every period after t (by assumption).
- Therefore, the *outcome* in every NE is that (D, D) is played in every period.
- The *strategies* chosen by each player may specify playing C in response to some history, but those histories will never actually occur.
- Outcomes and histories that do not occur in equilibrium are said to be *off the equilibrium path*.

Finitely repeated Prisoner's Dilemma

- What about SPNE? This is easier to prove: in the last subgame, the only NE is (D, D) regardless of the previous history. Going back one step, the only NE is (D, D), and so on, until we reach the beginning of the game. Therefore, the only SPNE is when both players' strategies is to play (D, D) after every history.
- Punishment cannot be sustained in the finitely repeated Prisoner's Dilemma because in the last period, there is no way to deter *Defect*.
- However, in an *infinitely* repeated game, there is *always* the possibility of punishment in the future.

NE of Repeated PD: Always D

- Suppose both players play *Always Defect*: they play D after any history.
- The sequence of outcomes will be $(D, D), (D, D), \ldots$
- The sequene of payoffs will be $(1, 1), (1, 1), \ldots$
- By our construction of the discounted average, this gives a discounted sum of 1 to both players.
- Is this a Nash equilibrium? Suppose Player 1 deviates to any strategy that does *not* result in the outcome sequence (D, D) in every period.
- In some period, Player 1 will get a payoff of 0 instead of 1.
- This must decrease Player 1's discounted sum, so there is no incentive to deviate.

NE of Repeated PD: Grim Trigger

- Recall the *Grim Trigger* strategy: Play C until the other player plays D, then punish by playing D forever.
- Suppose both players play *Grim Trigger*.
  - In the first period, both players C.
  - In the second period, no one has played D, so both players play C.
  - Same for period 3, 4, 5...
- The sequence of outcomes is: $(C, C), (C, C), \ldots$
- The sequence of payoffs for both players is $2, 2, \ldots$ with a discounted average of

$$(1-\delta)(2+\delta 2+\delta^2 2+\ldots) = (1-\delta)2\sum_{t=0}^{\infty}\delta^t = 2$$

NE of Repeated PD: Grim Trigger

- Now, suppose Player 2 deviates by playing some other strategy that actually results in a different sequence of outcomes.
- For the sequence of outcomes to be different, Player 2 must play D at least once.
- Then Player 1 will play D forever starting at $t+1$, Player 2's best response to this is to also play D forever starting at $t+1$.
- Player 2's sequence of payoffs starting at period t is $(3, 1, 1, 1...)$

$$(1-\delta)(3+\delta+\delta^2+\delta^3+...) = (1-\delta)(3+\frac{\delta}{1-\delta})$$

$$= 3(1-\delta)+\delta$$

- This deviation will give a higher payoff *Grim Trigger* (or any other strategy that results in an outcome where (C, C) is always played) if and only if:

$$3(1-\delta)+\delta > 2 \rightarrow \delta < \frac{1}{2}$$

NE of Repeated PD: Grim Trigger

$$3(1-\delta)+\delta > 2 \rightarrow \delta < \frac{1}{2}$$

- Therefore, if $\delta \geq \frac{1}{2}$, both players playing *Grim Trigger* is a Nash equilibrium.
- And in general, one player playing Grim Trigger and the other playing any strategy that results in (C, C) every period is a Nash equilibrium.
- Note what the condition on δ implies: if players are *patient* enough, i.e. they place a high enough value on future payoffs, then the threat of punishment is enough to deter *Defect* in the present.
- If players have a sufficiently low discount factor δ, then the short-term gain of playing D outweighs the long-term gain of avoiding punishment.

NE: Tit-for-Tat

- Recall $Tit - for - Tat$: play C at the beginning of the game, then play whatever the other player chose in the previous round.
- If both players play this strategy, the outcome will be (C, C) each period, with a payoff stream of $(2, 2, 2...)$.
- Suppose Player 2 plays another strategy that plays D at time t. Player 1 will therefore play D in $t + 1$.
- Player 2 can either:
  - revert to C at $t + 1$, in which case we are back in our original situation, or
  - play D in $t + 1$, which guarantees in Player 1 playing D again in $t + 2$.
- If one deviation in the original situation is optimal, then repeated deviation must also be optimal, since the game reverts back to the original situation after a single deviation.

NE: Tit-for-Tat

- In short, Player 2 can deviate in two ways:
  - Play a strategy that alternates between C and D. The outcomes will alternate between (C, D) and (D, C). This gives a payoff stream of $(3, 0, 3, 0, \ldots)$ with a discounted average of

$$(1-\delta)\frac{3}{1-\delta^2} = \frac{3}{1+\delta}$$

  - Play a strategy that plays D in each period. The outcomes will be (D, D) in each period starting from $t+1$. This gives a payoff stream of $(3, 1, 1, 1, \ldots)$ with a discounted average of

$$3(1-\delta) + \delta = 3 - 2\delta$$

NE: Tit-for-Tat

- Comparing the discounted averages, both players playing $Tit - for - Tat$ can be a NE when:

$$2 \geq \frac{3}{1+\delta} \text{ and } 2 \geq 3 - 2\delta$$

- which are both satisfied if $\delta \geq \frac{1}{2}$.

- So far, we have shown that some strategies, when played by both players, can be a Nash equilibrium.
- In general, there are an infinite number of possible strategies for repeated games, and therefore an infinite number of possible ways in which a player can deviate from a NE.
- This makes it difficult to prove whether any given pair of strategies is a NE.
- In contrast, when we consider subgame perfect NE, there are only a limited number of ways in which a player can deviate, which makes it much easier to show if a given pair of strategies is a SPNE.

Announcements

- Homework #4 is due today.
- Homework #5 will be posted on the web site later today, due in two weeks.