14.12 Game Theory Lecture Notes*
Lectures 7-9

Muhamet Yildiz

In these lectures we analyze dynamic games (with complete information). We first analyze the perfect information games, where each information set is singleton, and develop the notion of backwards induction. Then, considering more general dynamic games, we will introduce the concept of the subgame perfection. We explain these concepts on economic problems, most of which can be found in Gibbons.

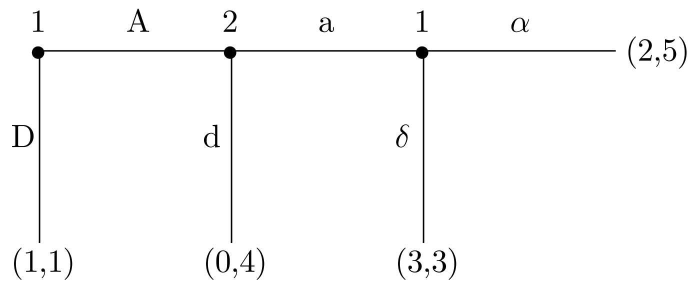
1 Backwards induction

The concept of backwards induction corresponds to the assumption that it is common knowledge that each player will act rationally at each node where he moves – even if his rationality would imply that such a node will not be reached.[1] Mechanically, it is computed as follows. Consider a finite horizon perfect information game. Consider any node that comes just before terminal nodes, that is, after each move stemming from this node, the game ends. If the player who moves at this node acts rationally, he will choose the best move for himself. Hence, we select one of the moves that give this player the highest payoff. Assigning the payoff vector associated with this move to the node at hand, we delete all the moves stemming from this node so that we have a shorter game, where our node is a terminal node. Repeat this procedure until we reach the origin.

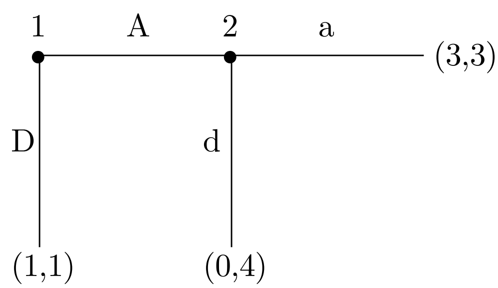
*These notes do not include all the topics that will be covered in the class. See the slides for a more complete picture.

[1] More precisely: at each node i the player is certain that all the players will act rationally at all nodes j that follow node i; and at each node i the player is certain that at each node j that follows node i the player who moves at j will be certain that all the players will act rationally at all nodes k that follow node j,...ad infinitum.

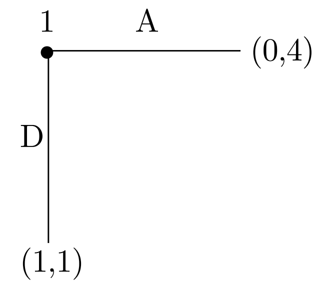
Example: consider the following well-known game, called as *the centipedes game*. This game illustrates the situation where it is mutually beneficial for all players to stay in a relation, while a player would like to exit the relation, if she knows that the other player will exit in the next day.



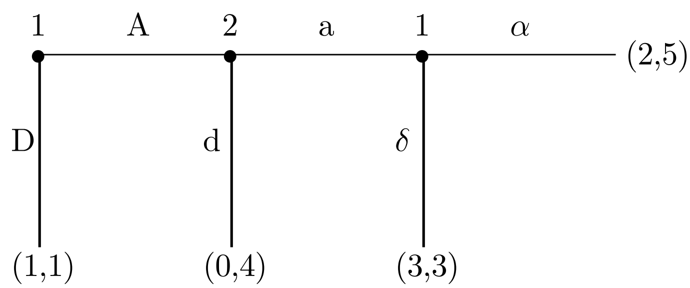
In the third day, player 1 moves, choosing between going across (α) or down (δ). If he goes across, he would get 2; if he goes down, he will get 3. Hence, we reckon that he will go down. Therefore, we reduce the game as follows:



In the second day, player 2 moves, choosing between going across (a) or down (d). If she goes across, she will get 3; if she goes down, she will get 4. Hence, we reckon that she will go down. Therefore, we reduce the game further as follows:



Now, player 1 gets 0 if he goes across (A), and gets 1 if he goes down (D). Therefore, he goes down. The equilibrium that we have constructed is as follows:



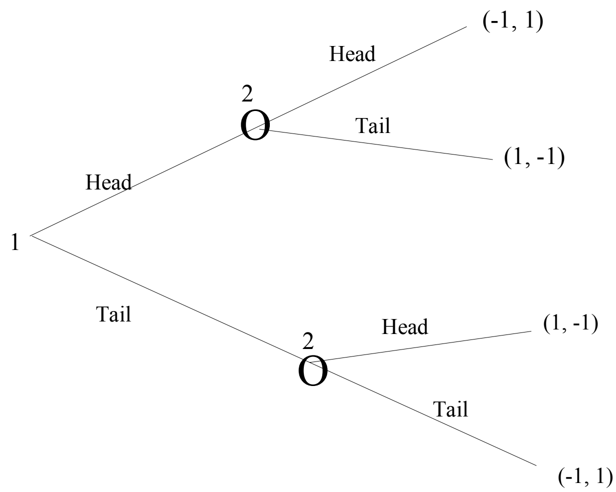
That is, at each node, the player who is to move goes down, exiting the relation.

Let's go over the assumptions that we have made in constructing our equilibrium. We assumed that player 1 will act rationally at the last date, when we reckoned that he goes down. When we reckoned that player 2 goes down in the second day, we assumed that player 2 assumes that player 1 will act rationally on the third day, and also assumed that she is rational, too. On the first day, player 1 anticipates all these. That is, he is assumed to know that player 2 is rational, and that she will keep believing that player 1 will act rationally on the third day.

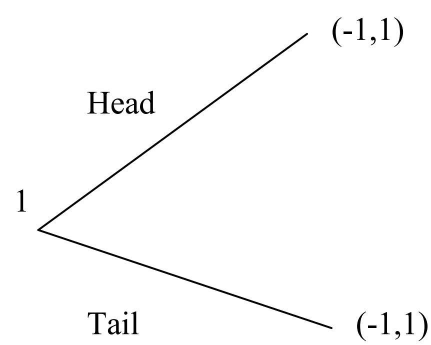
This example also illustrates another notion associated with backwards induction – commitment (or the lack of commitment). Note that the outcomes on the third day (i.e., (3,3) and (2,5)) are both strictly better than the equilibrium outcome (1,0). But they cannot reach these outcomes, because player 2 cannot commit to go across, whence player 1 exits the relation in the first day. There is also a further commitment problem in this example. If player 1 where able to commit to go across on the third day, player

2 would definitely go across on the second day, whence player 1 would go across on the first. Of course, player 1 cannot commit to go across on the third day, and the game ends in the first day, yielding the low payoffs (1,0).

As another example, let us apply backwards induction to the Matching Pennies with Perfect Information:



If player 1 chooses Head, player 2 will Head; and if 1 chooses Tail, player 2 will prefer Tail, too. Hence, the game is reduced to



In that case, Player 1 will be indifferent between Head and Tail, choosing any of these two option or any randomization between these two acts will give us an equilibrium with backwards induction.

At this point, you should stop and study the Stackelberg duopoly in Gibbons. You should also check that there is also a Nash equilibrium of this game in which the follower produces the Cournot quantity irrespective of what the leader produces, and the leader produces the Cournot quantity. Of course, this is not consistent with backwards induction: when the follower knows that the leader has produced the Stackelberg quantity, he will change his mind and produce a lower quantity, the quantity that is computed during the backwards induction. For this reason, we say that this Nash equilibrium is based on a non-credible threat (of the follower).

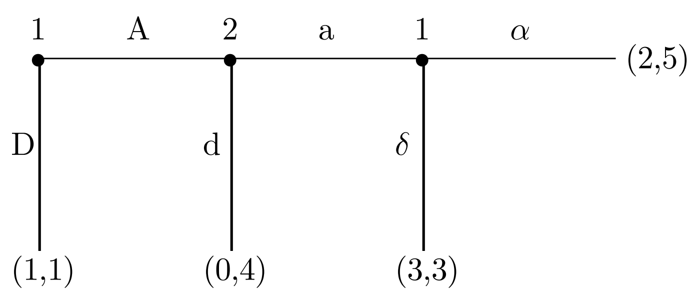
Backwards induction is a powerful solution concept with some intuitive appeal. Unfortunately, we cannot apply it beyond perfect information games with a finite horizon. Its intuition, however, can be extended beyond these games through subgame perfection.

2 Subgame perfection

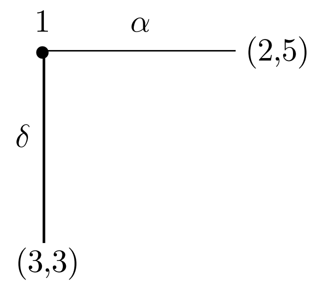
A main property of backwards induction is that, when we confine ourselves to a subgame of the game, the equilibrium computed using backwards induction remains to be an equilibrium (computed again via backwards induction) of the subgame. Subgame perfection generalizes this notion to general dynamic games:

Definition 1 *A Nash equilibrium is said to be* subgame perfect *if an only if it is a Nash equilibrium in every subgame of the game.*

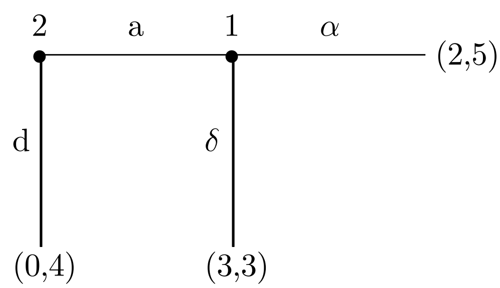
What is a subgame? In any given game, there may be some smaller games embedded; we call each such embedded game a subgame. Consider, for instance, the centipedes game (where the equilibrium is drawn in thick lines):



This game has three subgames. Here is one subgame:



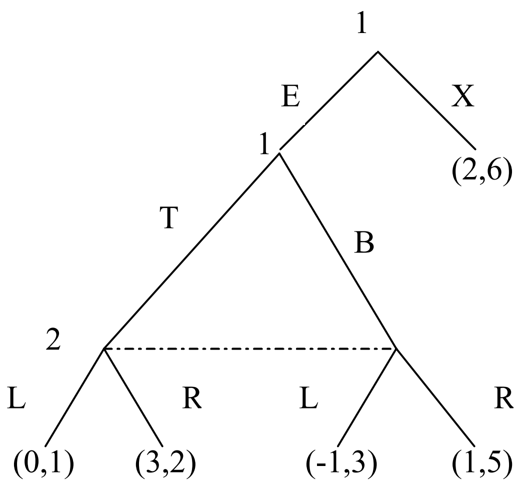
This is another subgame:



And the third subgame is the game itself. We call the first two subgames (excluding the game itself) proper. Note that, in each subgame, the equilibrium computed via backwards induction remains to be an equilibrium of the subgame.

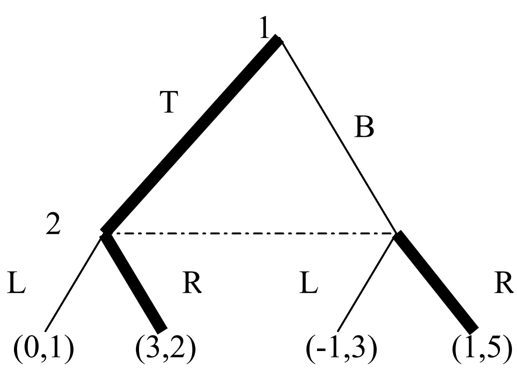
Now consider the matching penny game with perfect information. In this game, we have three subgames: one after player 1 chooses Head, one after player 1 chooses Tail, and the game itself. Again, the equilibrium computed through backwards induction is a Nash equilibrium at each subgame.

Now consider the following game.

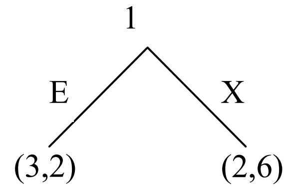


We cannot apply backwards induction in this game, because it is not a perfect information game. But we can compute the subgame perfect equilibrium. This game has two subgames: one starts after player 1 plays E; the second one is the game itself. We compute the subgame perfect equilibria as follows. We first compute a Nash equilibrium of the subgame, then fixing the equilibrium actions as they are (in this subgame), and taking the equilibrium payoffs in this subgame as the payoffs for entering in the subgame, we compute a Nash equilibrium in the remaining game.

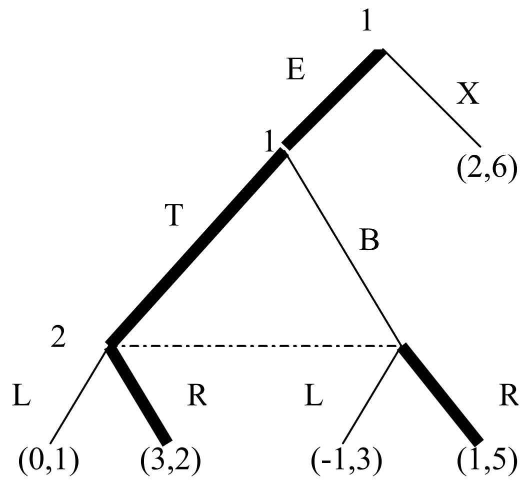
The subgame has only one Nash equilibrium, as T dominates B: Player 1 plays T and 2 plays R, yielding the payoff vector (3,2).



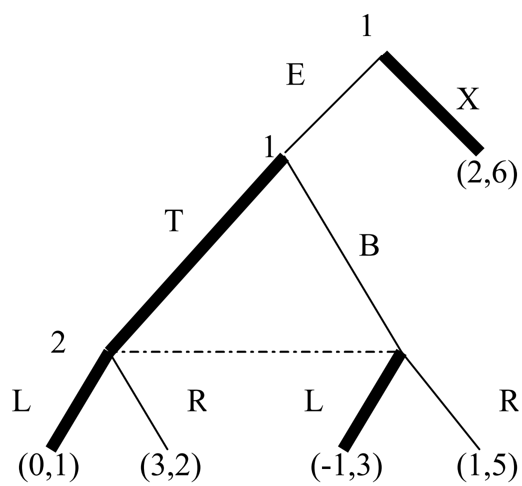
Given this, the remaining game is



where player 1 chooses E. Thus, the subgame-perfect equilibrium is as follows.

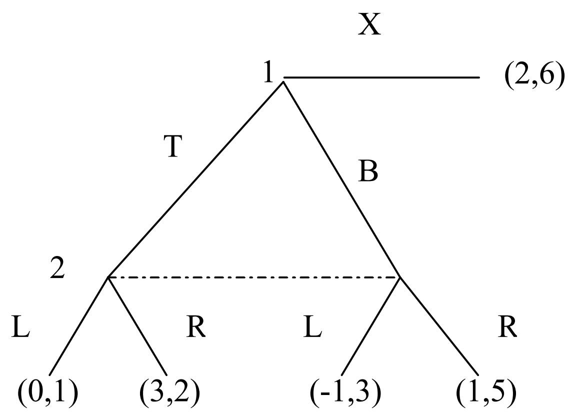


Note that there are other Nash Equilibria; one of them is depicted below.

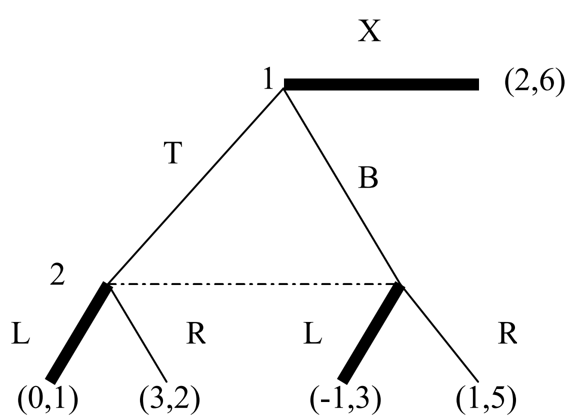


You should be able to check that this is a Nash equilibrium. But it is not subgame perfect, for, in the proper subgame, 2 plays a strictly dominated strategy.

Now, consider the following game, which is essentially the same game as above, with a slight difference that here player 1 makes his choices at once:



Note that the only subgame of this game is itself, hence any Nash equilibrium is subgame perfect. In particular, the non-subgame-perfect Nash equilibrium of the game above is subgame perfect. In the new game it takes the following form:



At this point you should stop reading and study "tariffs and imperfect international competition".

3 Sequential Bargaining

Imagine that two players own a dollar, which they can use only after they decide how to divide it. Each player is risk-neutral and discounts the future exponentially. That is, if a player gets x dollar at day t, his payoff is $\delta^t x$ for some $\delta \in (0,1)$. The set of all feasible divisions is $D = \{(x,y) \in [0,1]^2 \, | x + y \leq 1\}$. Consider the following scenario. In the first day player one makes an offer $(x_1, y_1) \in D$. Then, knowing what has been offered, player 2 accepts or rejects the offer. If he accepts the offer, the offer is implemented, yielding payoffs (x_1, y_1). If he rejects the offer, then they wait until the next day, when player 2 makes an offer $(x_2, y_2) \in D$. Now, knowing what player 2 has offered, player 1 accepts or rejects the offer. If player 1 accepts the offer, the offer is implemented, yielding payoffs $(\delta x_2, \delta y_2)$. If player two rejects the offer, then the game ends, when they lose the dollar and get payoffs (0,0).

Let us analyze this game. On the second day, if player 1 rejects the offer, he gets 0. Hence, he accepts any offer that gives him more than 0, and he is indifferent between accepting and rejecting any offer that gives him 0. Assume that he accepts the offer (0,1).[2] Then, player 2 would offer (0,1), which is the best player 2 can get. Therefore, if they do not agree on the first day, on the second day, player 2 takes the entire dollar, leaving player 1 nothing. The value of taking the dollar on the next day for player 2 is δ. Hence, on the first day, player 2 will accept any offer that gives him more than δ, will reject any offer that gives him less than δ, and he is indifferent between accepting and rejecting any offer that gives him δ. As above, assume that player 2 accepts the offer $(1-\delta, \delta)$. In that case, player 1 will offer $(1-\delta, \delta)$, which will be accepted. For any division that gives player 1 more than $1-\delta$ will give player 2 less than δ, and will be rejected.

Now, consider the game in which the game above is repeated n times. That is, if they have not yet reached an agreement by the end of the second day, on the third day, player 1 makes an offer $(x_3, y_3) \in D$. Then, knowing what has been offered, player 2 accepts or rejects the offer. If he accepts the offer, the offer is implemented, yielding payoffs $(\delta^2 x_3, \delta^2 y_3)$. If he rejects the offer, then they wait until the next day, when player 2 makes an offer $(x_4, y_4) \in D$. Now, knowing what player 2 has offered, player 1 accepts or rejects the offer. If player 1 accepts the offer, the offer is implemented, yielding payoffs $(\delta^3 x_4, \delta^3 y_4)$. If player two rejects the offer, then they go to the 5th day... And this goes on like this until the end of day $2n$. If they have not yet agreed at the end of that day,

[2] In fact, player 1 must accept (0,1) in equilibrium. For, if he doesn't accept (0,1), the best response of player 2 will be empty, inconsistent with an equilibrium. (Any offer $(\epsilon, 1-\epsilon)$ of player 2 will be accepted. But for any offer $(\epsilon, 1-\epsilon)$, there is a better offer $(\epsilon/2, 1-\epsilon/2)$, which will also be accepted.)

the game ends, when they lose the dollar and get payoffs (0,0).

The subgame perfect equilibrium will be as follows. At any day $t = 2n - 2k$ (k is a non-negative integer), player 1 accepts any offer (x, y) with

$$x \geq \frac{\delta\left(1-\delta^{2k}\right)}{1+\delta}$$

and will reject any offer (x, y) with

$$x < \frac{\delta\left(1-\delta^{2k}\right)}{1+\delta};$$

and player 2 offers

$$(x_t, y_t) = \left(\frac{\delta\left(1-\delta^{2k}\right)}{1+\delta}, 1 - \frac{\delta\left(1-\delta^{2k}\right)}{1+\delta}\right) \equiv \left(\frac{\delta\left(1-\delta^{2k}\right)}{1+\delta}, \frac{1+\delta^{2k+1}}{1+\delta}\right).$$

And at any day $t - 1 = 2n - 2k - 1$, player 2 accepts an offer (x, y) iff

$$y \geq \frac{\delta\left(1+\delta^{2k+1}\right)}{1+\delta};$$

and Player 1 will offer

$$(x_{t-1}, y_{t-1}) = \left(1 - \frac{\delta\left(1+\delta^{2k+1}\right)}{1+\delta}, \frac{\delta\left(1+\delta^{2k+1}\right)}{1+\delta}\right) \equiv \left(\frac{1-\delta^{2k+2}}{1+\delta}, \frac{\delta\left(1+\delta^{2k+1}\right)}{1+\delta}\right).$$

We can prove this is the equilibrium given by backwards induction using mathematical induction on k. (That is, we first prove that it is true for $k = 0$; then assuming that it is true for some $k - 1$, we prove that it is true for k.)

Proof. Note that for $k = 0$, we have the last two periods, identical to the 2-period example we analyzed above. Putting $k = 0$, we can easily check that the behavior described here is the same as the equilibrium behavior in the 2-period game. Now, assume that, for some $k - 1$ the equilibrium is as described above. That is, at the beginning of date $t + 1 := 2n - 2(k-1) - 1 = 2n - 2k + 1$, player 1 offers

$$(x_{t+1}, y_{t+1}) = \left(\frac{1-\delta^{2(k-1)+2}}{1+\delta}, \frac{\delta\left(1+\delta^{2(k-1)+1}\right)}{1+\delta}\right) = \left(\frac{1-\delta^{2k}}{1+\delta}, \frac{\delta\left(1+\delta^{2k-1}\right)}{1+\delta}\right);$$

and his offer is accepted. At date $t = 2n - 2k$, player one accepts an offer iff the offer is at least as good as having $\frac{1-\delta^{2k}}{1+\delta}$ in the next day, which is worth $\frac{\delta\left(1-\delta^{2k}\right)}{1+\delta}$. Therefore, he will accept an offer (x, y) iff

$$x \geq \frac{\delta\left(1-\delta^{2k}\right)}{1+\delta};$$

as we have described above. In that case, the best player 2 can do is to offer

$$(x_t, y_t) = \left(\frac{\delta\left(1-\delta^{2k}\right)}{1+\delta}, 1 - \frac{\delta\left(1-\delta^{2k}\right)}{1+\delta} \right) = \left(\frac{\delta\left(1-\delta^{2k}\right)}{1+\delta}, \frac{1+\delta^{2k+1}}{1+\delta} \right).$$

For any offer that gives 2 more than y_t will be rejected in which case player 2 will get

$$\delta y_{t+1} = \frac{\delta^2\left(1+\delta^{2k-1}\right)}{1+\delta} < y_t.$$

That is, at t player 2 offers (x_t, y_t); and it is accepted. In that case, at $t-1$, player 2 will accept an offer (x, y) iff

$$y \geq \delta y_t = \frac{\delta\left(1+\delta^{2k+1}\right)}{1+\delta}.$$

In that case, at $t-1$, player 1 will offer

$$(x_{t-1}, y_{t-1}) \equiv (1-\delta y_t, \delta y_t) = \left(\frac{1-\delta^{2k+2}}{1+\delta}, \frac{\delta\left(1+\delta^{2k+1}\right)}{1+\delta} \right),$$

completing the proof. ■

Now, let $n \to \infty$. At any odd date t, player 1 will offer

$$(x_t^\infty, y_t^\infty) = \lim_{k\to\infty} \left(\frac{1-\delta^{2k+2}}{1+\delta}, \frac{\delta\left(1+\delta^{2k+1}\right)}{1+\delta} \right) = \left(\frac{1}{1+\delta}, \frac{\delta}{1+\delta} \right);$$

and any even date t player 2 will offer

$$(x_t^\infty, y_t^\infty) = \lim_{k\to\infty} \left(\frac{\delta\left(1-\delta^{2k}\right)}{1+\delta}, \frac{1+\delta^{2k+1}}{1+\delta} \right) = \left(\frac{\delta}{1+\delta}, \frac{1}{1+\delta} \right);$$

and the offers are barely accepted.