Models in Microeconomic Theory

Martin J. Osborne
Ariel Rubinstein

https://www.openbookpublishers.com

ISBN Paperback: 978-1-78374-892-1
ISBN Hardback: 978-1-78374-893-8
ISBN Digital (PDF): 978-1-78374-894-5
DOI: 10.11647/OBP.0211

This book is available in two versions, one that uses feminine pronouns and one that uses masculine pronouns. This version uses feminine pronouns.

Version: 2020.12.21 (s)

16 Extensive games

A market is currently served by a single incumbent. A competitor is considering entering the market. The incumbent wants to remain alone in the market and thus wishes to deter the competitor from entering. If the competitor enters, the incumbent can start a price war or can act cooperatively. A price war is the worst outcome for both parties; cooperative behavior by the incumbent is best for the competitor, and for the incumbent is better than a price war but worse than the competitor's staying out of the market.

We can model this situation as a strategic game. The competitor (player 1) decides whether to enter the market (*In*) or not (*Out*). If the competitor enters, the incumbent (player 2) decides whether to *Fight* the competitor or to *Cooperate* with it. The following table shows the game.

	Fight	*Cooperate*
In	0,0	2,2
Out	1,5	1,5

The game has two pure Nash equilibria, (*In*, *Cooperate*) and (*Out*, *Fight*). In the second equilibrium the incumbent plans to fight the competitor if she enters, a decision that deters the competitor from entering.

The formulation of the situation as this strategic game makes sense if the incumbent can decide initially to fight a competitor who enters the market and cannot reconsider this decision if the competitor does in fact enter. If the incumbent can reconsider her decision, the analysis is less reasonable: after the competitor enters, the incumbent is better off being cooperative than waging a price war. In this case, a model in which the timing of the decisions is described explicitly is more suitable for analyzing the situation. One such model is illustrated in Figure 16.1. Play starts at the initial node, indicated in the figure by a small circle. The label above this node indicates the player whose move starts the game (player 1, the competitor). The branches emanating from the node, labeled *In* and *Out*, represent the actions available to the competitor at the start of the game. If she chooses *Out*, the game is over. If she chooses *In*, player 2, the incumbent, chooses between *Cooperate* and *Fight*. The payoffs at the endpoints represent the players' preferences: player 1 (whose payoff is listed first in each pair) prefers (*In*, *Cooperate*) to *Out* to (*In*, *Fight*), and player 2 prefers *Out* to (*In*, *Cooperate*) to (*In*, *Fight*).

Chapter of *Models in Microeconomic Theory* by Martin J. Osborne and Ariel Rubinstein. Version 2020.12.21 (s).
 https://doi.org/10.11647/OBP.0211.16

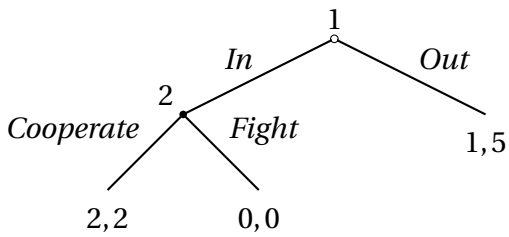


Figure 16.1 The entry game described in the introduction to the chapter.

We refer to each sequence of actions as a history. In Figure 16.1 there are five histories. The initial node represents the null history: no action has yet been chosen. The node shown by a small disk represents the history (*In*). Each of the three other histories, (*In*, *Cooperate*), (*In*, *Fight*), and (*Out*), leads to an endpoint of the game. We refer to these histories as terminal, and to the other histories, after which a player has to choose an action, as nonterminal.

16.1 Extensive games and subgame perfect equilibrium

An extensive game is specified by a set of players, a set of possible histories, a player function, which assigns a player to each nonterminal history, and the players' preferences over the terminal histories. We focus on games in which every history is finite.

Definition 16.1: Finite horizon extensive game

A (*finite horizon*) *extensive game* $\langle N, H, P, (\succsim^i)_{i \in N} \rangle$ has the following components.

Players

A set of players $N = \{1, \ldots, n\}$.

Histories

A set H of *histories*, each of which is a finite sequence of *actions*. The empty history, $\varnothing$, is in H, and if $(a_1, a_2, \ldots, a_t) \in H$ then also $(a_1, a_2, \ldots, a_{t-1}) \in H$.

A history $h \in H$ is *terminal* if there is no x such that $(h, x) \in H$. The set of terminal histories is denoted Z. (We use the notation $(h, a_1, \ldots, a_t)$ for the history that starts with the history h and continues with the actions $a_1, \ldots, a_t$).

Player function

A function $P : H \setminus Z \to N$, the *player function*, which assigns a player to each nonterminal history (the player who moves after the history).

Preferences

For each player $i \in N$, a preference relation $\succsim^i$ over Z.

We interpret this model as capturing a situation in which every player, when choosing an action, knows all actions previously chosen. For this reason, the model is usually called an extensive game *with perfect information*. A more general model, which we do not discuss, allows the players to be imperfectly informed about the actions previously chosen.

The example in the introduction, represented in Figure 16.1, is the extensive game $\langle N, H, P, (\succsim^i)_{i \in N} \rangle$ in which

- $N = \{1, 2\}$
- $H = \{\emptyset, (Out), (In), (In, Cooperate), (In, Fight)\}$ (with $Z = \{(Out), (In, Cooperate), (In, Fight)\}$)
- $P(\emptyset) = 1$ and $P(In) = 2$
- $(In, Cooperate) \succ^1 (Out) \succ^1 (In, Fight)$ and $(Out) \succ^2 (In, Cooperate) \succ^2 (In, Fight)$.

Notice that we use the notation $P(In)$ instead of $P((In))$; later we similarly write $P(a_1, \ldots, a_t)$ instead of $P((a_1, \ldots, a_t))$.

A key concept in the analysis of an extensive game is that of a strategy. A player's strategy is a specification of an action for every history after which the player has to move.

Definition 16.2: Strategy in extensive game

A *strategy* of player $i \in N$ in the extensive game $\langle N, H, P, (\succsim^i)_{i \in N} \rangle$ is a function that assigns to every history $h \in H \setminus Z$ for which $P(h) = i$ an action in $\{x : (h, x) \in H\}$, the set of actions available to her after h.

A key word in this definition is "every": a player's strategy specifies the action she chooses for *every* history after which she moves, even histories that do not occur if she follows her strategy. For example, in the game in Figure 16.2, one strategy of player 1 is s^1 with $s^1(\emptyset) = A$ and $s^1(B, G) = I$. This strategy specifies the action of player 1 after the history (B, G) although this history does not occur if player 1 uses s^1 and hence chooses A at the start of the game. Thus the notion of a strategy does not correspond to the notion of a strategy in everyday language. We discuss this issue further in Section 16.2.

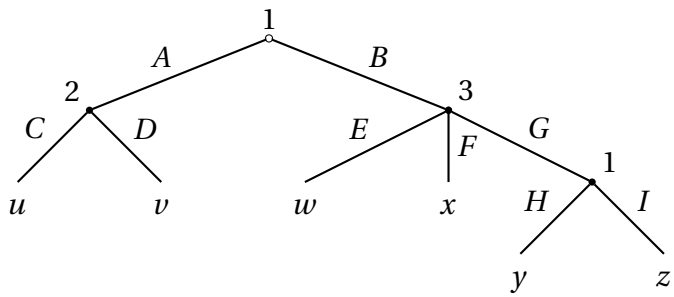


Figure 16.2 An example of an extensive game.

Each strategy profile generates a unique terminal history $(a_1 \ldots, a_T)$ as the players carry out their strategies. The first component of this history, a_1, is the action $s^{P(\varnothing)}(\varnothing)$ specified by the strategy $s^{P(\varnothing)}$ of player $P(\varnothing)$, who moves at the start of the game. This action determines the player who moves next, $P((a_1))$; her strategy $s^{P(a_1)}$ determines the next action, $a_2 = s^{P(a_1)}(a_1)$, and so forth.

Definition 16.3: Terminal history generated by strategy profile

Let s be a strategy profile for the extensive game $\langle N, H, P, (\succsim^i)_{i \in N} \rangle$. The *terminal history generated by* s is $(a_1, \ldots, a_T)$ where $a_1 = s^{P(\varnothing)}(\varnothing)$ and $a_{t+1} = s^{P(a_1,\ldots,a_t)}(a_1, \ldots, a_t)$ for $t = 1, \ldots, T-1$.

The main solution concept we use for extensive games is subgame perfect equilibrium. Before defining this notion, we define a Nash equilibrium of an extensive game: a strategy profile with the property that no player can induce a more desirable outcome for herself by deviating to a different strategy, given the other players' strategies.

Definition 16.4: Nash equilibrium of extensive game

Let $\Gamma = \langle N, H, P, (\succsim^i)_{i \in N} \rangle$ be an extensive game. A strategy profile s is a *Nash equilibrium* of Γ if for every player $i \in N$ we have

$$z(s) \succsim^i z(s^{-i}, r^i) \text{ for every strategy } r^i \text{ of player } i,$$

where, for any strategy profile σ, $z(\sigma)$ is the terminal history generated by σ.

The entry game, given in Figure 16.1, has two Nash equilibria: (*In*, *Cooperate*) and (*Out*, *Fight*). The latter strategy pair is a Nash equilibrium because given the incumbent's strategy *Fight*, the strategy *Out* is optimal for the competitor, and given the competitor's strategy *Out*, the strategy *Fight* is optimal for the

incumbent. In fact, if the competitor chooses *Out*, then any strategy for the incumbent is optimal.

The non-optimality of *Fight* for the incumbent if the competitor chooses *In* does not interfere with the status of (*Out*, *Fight*) as a Nash equilibrium: the notion of Nash equilibrium considers the optimality of a player's strategy only at the start of the game, before any actions have been taken.

The notion of subgame perfect equilibrium, by contrast, requires that each player's strategy is optimal, given the other players' strategies, after every possible history, whether or not the history occurs if the players follow their strategies. To define this notion, we first define, for any strategy profile s and nonterminal history h, the outcome (terminal history) that is reached if h occurs and then the players choose the actions specified by s.

Definition 16.5: Terminal history extending history

Let s be a strategy profile for the extensive game $\langle N, H, P, (\succsim^i)_{i \in N} \rangle$ and let h be a nonterminal history. The *terminal history extending h generated by s*, denoted $z(h, s)$, is $(h, a_1, \ldots, a_T)$ where $a_1 = s^{P(h)}(h)$ and $a_{t+1} = s^{P(h, a_1, \ldots, a_t)}(h, a_1, \ldots, a_t)$ for $t = 1, \ldots, T-1$.

In the game in Figure 16.2, for example, if $h = B$ and the players' strategies specify $s^1(\varnothing) = A$, $s^1(B, G) = H$, $s^2(A) = C$, and $s^3(B) = G$, then the terminal history extending h generated by s is (B, G, H).

Definition 16.6: Subgame perfect equilibrium of extensive game

Let $\Gamma = \langle N, H, P, (\succsim^i)_{i \in N} \rangle$ be an extensive game. A strategy profile $s = (s^i)_{i \in N}$ is a *subgame perfect equilibrium* of Γ if for every player $i \in N$ and every nonterminal history h for which $P(h) = i$ we have

$$z(h, s) \succsim^i z(h, (s^{-i}, r^i)) \text{ for every strategy } r^i \text{ of player } i,$$

where, for any history h and strategy profile σ, $z(h, \sigma)$ is the terminal history extending h generated by σ.

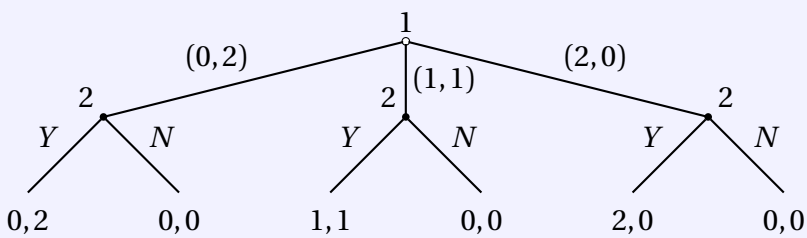
The difference between this definition and that of a Nash equilibrium is the phrase "and every nonterminal history h for which $P(h) = i$". The notion of Nash equilibrium requires that each player's strategy is optimal at the beginning of the game (given the other players' strategies) whereas the notion of subgame perfect equilibrium requires that it is optimal after every history (given the other players' strategies), even ones that are not consistent with the strategy profile.

Every subgame perfect equilibrium is a Nash equilibrium, but some Nash equilibria are not subgame perfect equilibria. In a subgame perfect equilibrium of the entry game (Figure 16.1), the incumbent's strategy must specify *Cooperate* after the history *In*, because the incumbent prefers the terminal history (*In*, *Cooperate*) to the terminal history (*In*, *Fight*). Given this strategy of the incumbent, the competitor's best strategy is *In*. The Nash equilibrium (*Out*, *Fight*) is not a subgame perfect equilibrium because *Fight* is not optimal for the incumbent after the history *In*.

Example 16.1: Ultimatum game

Two players have to agree how to allocate two indivisible units of a good between themselves. If they do not agree then each of them gets nothing. They use the take-it-or-leave-it protocol: Player 1 proposes one of the three partitions of the two units, which player 2 either accepts or rejects. Each player cares only about the number of units of the good she gets (the more the better) and not about the number of units the other player gets.

Denote by (x^1, x^2) the proposal in which i gets x^i, with $x^1 + x^2 = 2$. The situation is modeled by the following extensive game. At the start of the game (the null history, $\varnothing$), player 1 makes one of the three proposals, $(2,0)$, $(1,1)$, and $(0,2)$, and after each of these proposals player 2 either agrees (Y) or disagrees (N).



Player 1 has three strategies and player 2 has eight. Each of player 2's strategies specifies her reaction to each possible proposal of player 1; examples are (Y, Y, Y), in which she accepts all proposals, and (Y, N, N), in which she accepts the proposal $(0,2)$ and rejects the two other proposals.

The game has several Nash equilibria. In particular, for *any* allocation the game has a Nash equilibrium with that outcome: player 1 proposes the allocation and player 2 accepts that allocation and rejects the other two. The strategy pair $((2,0),(N,N,N))$ is also a Nash equilibrium, which yields disagreement.

Consider the Nash equilibrium $((0,2),(Y,N,N))$. Player 2's strategy accepts only the offer $(0,2)$, which gives her both units. However, her threat

to reject $(1,1)$ is not credible, because if player 1 proposes that allocation, player 2 prefers to accept it and get one unit than to reject it and get nothing.

In any subgame perfect equilibrium, player 2's action after every proposal of player 1 must be optimal, so that she accepts the proposals $(0,2)$ and $(1,1)$. She is indifferent between accepting and rejecting the proposal $(2,0)$, so either action is possible in a subgame perfect equilibrium. Thus the only strategies of player 2 consistent with subgame perfect equilibrium are (Y,Y,Y) and (Y,Y,N). Player 1 optimally proposes $(2,0)$ if player 2 uses the first strategy, and $(1,1)$ if she uses the second strategy. Hence the game has two subgame perfect equilibria, $((2,0),(Y,Y,Y))$ and $((1,1),(Y,Y,N))$.

If there are K units of the good to allocate, rather than two, then also the game has two subgame perfect equilibria, $((K,0),(Y,\ldots,Y))$ and $((K-1,1),(Y,\ldots,Y,N))$. In the first equilibrium player 1 proposes that she gets all K units and player 2 agrees to all proposals. In the second equilibrium player 2 plans to reject only the proposal that gives him no units and player 1 proposes that player 2 gets exactly one unit.

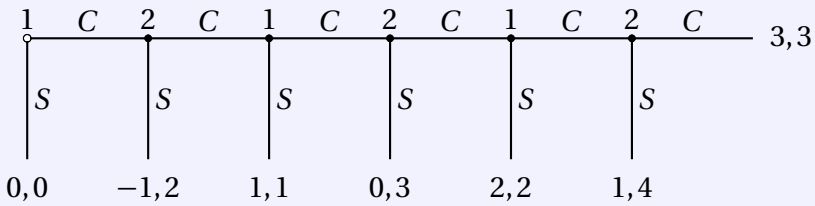
Example 16.2: Centipede game

Two players, 1 and 2, alternately have the opportunity to stop their interaction, starting with player 1; each player has T opportunities to do so. Whenever a player choose to continue (C), she loses \$1 and the other player gains \$2. Each player aims to maximize the amount of money she has at the end of the game.

This situation may be modeled as an extensive game in which the set of histories consists of $2T$ nonterminal histories of the form $C_t = (C,\ldots,C)$, where $t \in \{0,\ldots,2T-1\}$ is the number of occurrences of C ($C_0 = \varnothing$, the null history), and $T+1$ terminal histories, C_{2T} (both players always choose C) and $S_t = (C,\ldots,C,S)$ for $t \in \{0,\ldots,2T-1\}$, where t is the number of occurrences of C (the players choose C in the first t periods and then one of them chooses S).

After the history C_t, player 1 moves if t is even (including 0) and player 2 moves if t is odd. Each player's payoff is calculated by starting at 0, subtracting 1 whenever the player chooses C, and adding 2 whenever the other player chooses C. The diagram on the next page shows the game for $T=3$. (The shape of the tree is the reason for the name "centipede".)

Any pair of strategies in which each player plans to stop the game at the first opportunity is a Nash equilibrium. Given player 2's plan, player 1 can



only lose by changing her strategy, and given that player 1 intends to stop the game immediately, player 2 is indifferent between all her strategies.

In fact, we now show that in every Nash equilibrium player 1 stops the game immediately. That is, the only terminal history generated by a Nash equilibrium is S_0. For any pair of strategies that generates the terminal history S_t with $t \geq 1$, the player who moves after the history C_{t-1} can increase her payoff by changing her strategy to one that stops after this history, saving her the loss of continuing at this history. The terminal history C_{2T} occurs only if each player uses the strategy in which she plays C at every opportunity, in which case player 2 can increase her payoff by deviating to the strategy of stopping only at C_{2T-1}.

Although the outcome of every Nash equilibrium is S_0, the game has many Nash equilibria. In every equilibrium player 1 chooses S at the start of the game and player 2 chooses S after the history (C), but after longer histories each player's strategy may choose either C or S.

However, the game has a unique subgame perfect equilibrium, in which each player chooses S whenever she moves. The argument is by induction, starting at the end of the game: after the history C_{2T-1}, player 2 optimally stops the game, and if the player who moves after the history C_t for $t \geq 1$ stops the game, then the player who moves after C_{t-1} optimally does so.

When people play the game in experiments, they tend not to stop it immediately. There seem to be two reasons for the divergence from equilibrium. First, many people appear to be embarrassed by stopping the game to gain \$1 while causing the other player to lose \$2 when there is an opportunity for a large mutual gain. Second, people seem to continue at least for a while because they are not sure of their opponent's strategic reasoning, and given the potential gain they are ready to sacrifice \$1 to check her intentions.

16.2 What is a strategy?

Consider player 1's strategy (S, C, C) in the centipede game with $T = 3$. According to this strategy, player 1 plans to stop the game immediately, but plans to

continue at her later moves (after the histories (C, C) and (C, C, C, C)). To be a complete plan of action, player 1's strategy has to specify a response to every possible action of player 2. But a strategy in an extensive game does more than that. Under the strategy (S, C, C) player 1 plans to stop the game immediately, but specifies also her action in the event she has a second opportunity to stop the game, an opportunity that does not occur if she follows her own strategy and stops the game immediately. That is, the strategy specifies plans after contingencies that are inconsistent with the strategy.

In this respect the notion of a strategy in an extensive game does not correspond to a plan of action, which naturally includes actions only after histories consistent with the plan. In the centipede game with $T = 3$, player 1 has four natural plans of action: always continue, and stop at the tth opportunity for $t = 1$, 2, and 3.

Why do we define a strategy more elaborately than a plan of action? When player 2 plans her action after the history (C) she needs to think about what will happen if she does not stop the game. That is, she needs to think about the action player 1 will take after the history (C, C). The second component of player 1's strategy (S, C, C), which specifies an action after the history (C, C), can be thought of as player 2's belief about the action that player 1 will take after (C, C) if player 1 does not stop the game. Thus a pair of strategies in the centipede game, and in other extensive games in which players move more than once, is more than a pair of plans of action; it embodies also an analysis of the game that contains the beliefs of the players about what would happen after any history.

16.3 Backward induction

Backward induction is a procedure for selecting strategy profiles in an extensive game. It is based on the assumption that whenever a player moves and has a clear conjecture about what will happen subsequently, she chooses an action that leads to her highest payoff. The procedure starts by considering histories that are one action away from being terminal, and then works back one step at a time to the start of the game.

To describe the procedure, we first define the diameter of a history h to be number of steps remaining until the end of the game in the longest history that starts with h.

Definition 16.7: Diameter of history

The *diameter* of the history h in an extensive game is the largest number K for which there are actions $a_1, \ldots, a_K$ such that $(h, a_1, \ldots, a_K)$ is a history.

Note that the diameter of a history is zero if and only if the history is terminal, and

the diameter of the null history is the number of actions in the longest history in the game.

The backward induction procedure starts by specifying the action chosen by each player who moves after a history with diameter 1, and then works back in steps to the start of the game. As it does so, it associates with every history h the terminal history $z(h)$ that occurs if the game reaches h and then the players take the actions specified in the previous steps.

In the first step we define $z(h) = h$ for every terminal history h.

In the second step we consider histories with diameter 1. Let h be such a history, so that one action remains to be taken after h, by player $P(h)$. (In the game in Figure 16.3, the two such histories are A and (B,G), with $P(A) = 3$ and $P(B,G) = 1$.) For every action a of $P(h)$ after h, the history (h,a) has diameter 0 (i.e. it is terminal), and hence $z(h,a)$ is defined from the first step (it is equal to (h,a)). From among these actions, let $a = a^*(h)$ be one that maximizes $P(h)$'s payoff over all terminal histories $z(h,a)$, and set $s^{P(h)}(h) = a^*(h)$. Note that if there is more than one such action, we select one of them arbitrarily. (In the game in Figure 16.3, C is such an action for the history (A), and both H and I are such actions for the history (B,G). Either of these actions can be chosen at this step.) Define $z(h) = z(h, a^*(h))$, the terminal history that occurs if the game reaches h and then player $P(h)$ chooses $a^*(h)$.

The procedure continues working backwards until the start of the game. After step k, for every history h with diameter at most k an action for the player who moves after h is defined, together with the resulting terminal history $z(h)$, so that at step $k+1$, for every history with diameter $k+1$, we can find an optimal action for the player who moves after this history. At the end of the process, a strategy for each player in the game is defined.

Procedure: Backward induction

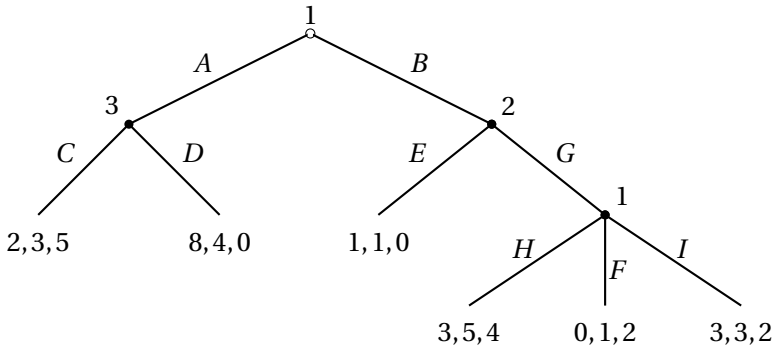
The *backward induction procedure* for an extensive game $\langle N, H, P, (\succsim^i)_{i \in N} \rangle$ generates a strategy profile s as follows. For any history $h \in H$, denote the diameter of h by $d(h)$.

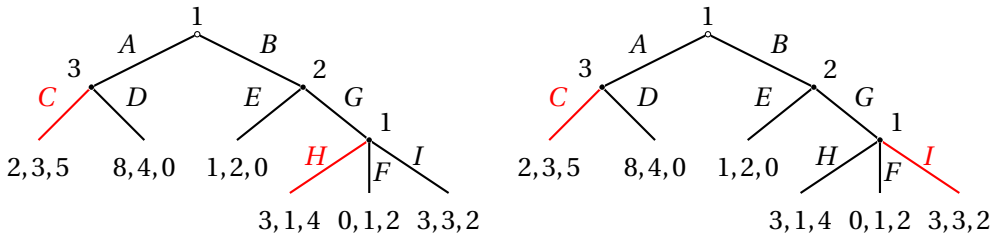
Initialization

For each history h with $d(h) = 0$ (that is, each terminal history), let $z(h) = h$.

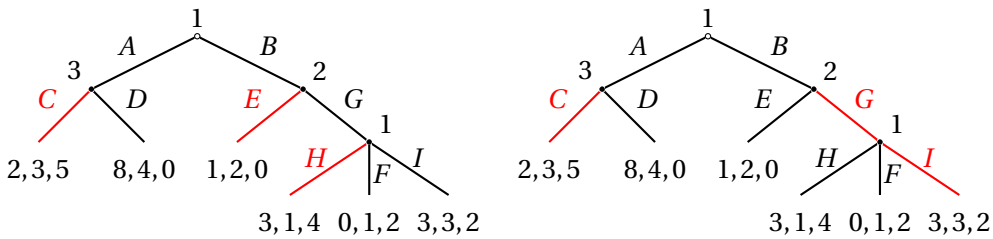
Inductive step

Assume that the terminal history $z(h)$ is defined for all $h \in H$ with $d(h) \in \{0, \ldots, k\}$ and $s^{P(h)}(h)$ is defined for all $h \in H$ with $d(h) \in \{1, \ldots, k\}$, where $k < d(\varnothing)$. For each history h with $d(h) = k+1$, let $a^*(h)$ be an action a that is best according to $P(h)$'s preferences over terminal histories $z(h,a)$, and set $s^{P(h)}(h) = a^*(h)$ and $z(h) = z(h, a^*(h))$.

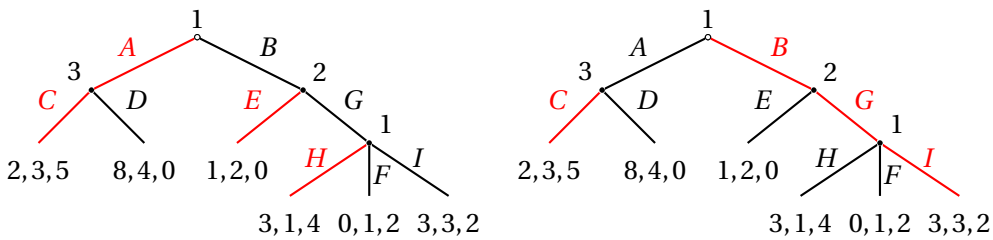




Step 1: Choosing optimal actions for the player who moves after each history with diameter 1. The action C is optimal after the history (A). Both H and I are optimal after the history (B,G). The diagrams show the resulting two possible outcomes of the step.



Step 2: Choosing optimal actions for the player who moves after the single history with diameter 2.



Step 3: Choosing optimal actions for the player who moves after the single history with diameter 3 (the initial history).

Figure 16.3 An example of backward induction. For this game, the procedure selects one of two strategy profiles, which yield one of the two terminal histories (A,C) and (B,G,I).

We say that the strategy profile s *is generated by backward induction* if for some choice of an optimal action after each history, this procedure generates s.

The procedure is well-defined only if an optimal action exists whenever the procedure calls for such an action. In particular it is well-defined for any game with a finite number of histories. If the number of actions after some history is not finite, an optimal action may not exist, in which case the procedure is not well-defined.

We now show that any strategy profile generated by the backward induction procedure is a subgame perfect equilibrium. To do so, we first give an alternative characterization of a subgame perfect equilibrium of an extensive game.

Recall that a strategy profile s is a subgame perfect equilibrium if after no history h does any player have a strategy that leads to a terminal history she prefers to the terminal history generated by s after h. In particular, for any history, the player who moves cannot induce an outcome better for her by changing only her action after that history, *keeping the remainder of her strategy fixed*. We say that a strategy profile with this property satisfies the one-deviation property.

Definition 16.8: One-deviation property of strategy profile

Let $\Gamma = \langle N, H, P, (\succsim^i)_{i \in N} \rangle$ be an extensive game. A strategy profile s for Γ satisfies the *one-deviation property* if for every player $i \in N$ and every nonterminal history $h \in H \setminus Z$ for which $P(h) = i$ we have

$$z(h,s) \succsim^i z(h,(s^{-i}, r^i)) \text{ for every strategy } r^i \text{ of player } i$$
$$\text{that differs from } s^i \text{ only in the action it specifies after } h.$$

A profile of strategies that is a subgame perfect equilibrium satisfies the one-deviation property. The reason is that a subgame perfect equilibrium requires, for any history and any player, that the player's strategy is optimal at that history among all strategies, whereas the one deviation property requires the optimality to hold only among the strategies that differ in the action planned after that history.

We now show that the converse is true: any strategy profile satisfying the one-deviation property is a subgame perfect equilibrium. To illustrate the argument suppose, to the contrary, that the strategy profile s satisfies the one-deviation property, generating the payoff u^i for some player i, but that after some history h at which i moves, i can obtain the payoff $v^i > u^i$ by changing the action specified by her strategy at both h, from say a to a', and at some history h' that extends h,

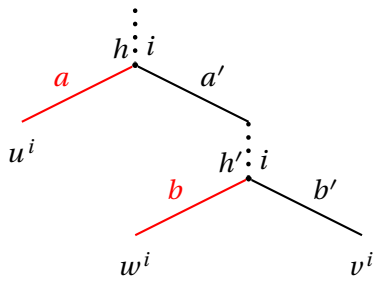


Figure 16.4 An illustration of the argument that a strategy profile that satisfies the one-deviation property is a subgame perfect equilibrium.

from say b to b' (given the other players' strategies). (See Figure 16.4.) Because s satisfies the one-deviation property, the payoff s generates for i starting at h' (after i changes her action only at h), say w^i, is at most u^i. But then $v^i > w^i$ and hence at h' player i can induce a higher payoff than w^i by changing only her action at h' from b to b' holding the rest of her strategy fixed, contradicting the assumption that s satisfies the one-deviation property.

Proposition 16.1: One-deviation property and SPE

For an extensive game (in which every terminal history is finite) a strategy profile satisfies the one-deviation property if and only if it is a subgame perfect equilibrium.

Proof

As we explained earlier, if a strategy profile is a subgame perfect equilibrium then it satisfies the one-deviation property.

Let s be a strategy profile that satisfies the one-deviation property. Assume, contrary to the claim, that s is not a subgame perfect equilibrium. Then for some player i there is a nonempty set H^* of histories such that after each $h \in H^*$, a change in player i's strategy generates a terminal history that she prefers to $z(h,s)$ (the terminal history extending h generated by s). For each $h \in H^*$, let $n(h)$ be the minimal number of histories h' after which i's strategy needs to specify an action different from $s^i(h')$ for the new strategy profile to generate a terminal history that i prefers to $z(h,s)$.

Let H^{**} be the subset of H^* for which $n(h)$ is minimal over $h \in H^*$. Let $h^* \in H^{**}$ be a longest history in H^{**} and let r^i be a strategy with $n(h^*)$ changes for which $z(h^*,(s^{-i},r^i)) \succ^i z(h^*,s)$. We have $r^i(h^*) \neq s^i(h^*)$ since otherwise, we have $(h^*,s^i(h^*)) \in H^{**}$, a history that is longer than h^*. The strategy r^i involves fewer than $n(h^*)$ changes after $(h^*,r^i(h^*))$ and thus, by

the minimality of $n(h^*)$,

$$z(h^*,(s^{-i},r^i)) = z((h^*,r^i(h^*)),(s^{-i},r^i)) \preccurlyeq^i z((h^*,r^i(h^*)),s).$$

Because s satisfies the one-deviation property, $z((h^*,r^i(h^*)),s) \preccurlyeq^i z(h^*,s)$, which implies that $z(h^*,(s^{-i},r^i) \preccurlyeq^i z(h^*,s)$, a contradiction.

Note that the proof uses the assumption that all histories are finite, and indeed if not all histories are finite, a strategy profile may satisfy the one deviation property and not be a subgame perfect equilibrium (see Problem 9).

In many games, this result greatly simplifies the verification that a strategy profile is a subgame perfect equilibrium, because it says that we need to check only whether, for each history, the player who moves can increase her payoff by switching to a different action after that history.

We now show that any strategy profile generated by the procedure of backward induction is a subgame perfect equilibrium, by arguing that it satisfies the one-deviation property.

Proposition 16.2: Backward induction and SPE

For an extensive game (in which every terminal history is finite), a strategy profile is generated by the backward induction procedure if and only if it is a subgame perfect equilibrium.

Proof

A strategy profile generated by the backward induction procedure by construction satisfies the one-deviation property. Thus by Proposition 16.1 it is a subgame perfect equilibrium.

Conversely, if a strategy profile is a subgame perfect equilibrium then it satisfies the one-deviation property, and hence is generated by the backward induction procedure where at each step we choose the actions given by the strategy profile.

An immediate implication of this result is that every extensive game with a finite number of histories has a subgame perfect equilibrium, because for every such game the backward induction procedure is well-defined.

Proposition 16.3: Existence of SPE in finite game

Every extensive game with a finite number of histories has a subgame perfect equilibrium.

Chess is an example of a finite extensive game. In the game, two players move alternately. The terminal histories are of three types: player 1 wins, player 2 wins, and the players draw. Each player prefers to win than to draw than to lose. The game is finite because once a position is repeated three times, a draw is declared. Although the number of histories is finite, it is huge, and currently no computer can carry out the backwards induction procedure for the game. However, we know from Proposition 16.3 that chess has a subgame perfect equilibrium. Modeled as a strategic game, chess is strictly competitive, so we know also (Proposition 15.3) that the payoffs in all Nash equilibria are the same and the Nash equilibrium strategies are maxmin strategies: either one of the players has a strategy that guarantees she wins, or each player has a strategy that guarantees the outcome is at least a draw.

Ticktacktoe is another example of a finite extensive game that is strictly competitive. For ticktacktoe, we know that each player can guarantee a draw. Chess is more interesting than ticktacktoe because whether a player can guarantee a win or a draw in chess is not known; the outcome of a play of chess depends on the player's cognitive abilities more than the outcome of a play of ticktacktoe. Models of bounded rationality, which we do not discuss in this book, attempt to explore the implications of such differences in ability.

16.4 Bargaining

This section presents several models of bargaining, and in doing so illustrates how an extensive game may be used to analyze an economic situation. Bargaining is a typical economic situation, as it involves a mixture of common and conflicting interests. The parties have a common interest in reaching an agreement, but differ in their evaluations of the possible agreements. Bargaining models are key components of economic models of markets in which exchange occurs through pairwise matches and the terms of exchange are negotiated. These market models differ from the market models presented in Part II of the book in that the individuals do not perceive prices as given.

For simplicity we confine ourselves to the case in which two parties, 1 and 2, bargain over the partition of a desirable pie of size 1. The set of possible agreements is

$$X = \{(x^1, x^2) : x^1 + x^2 = 1 \text{ and } x^i \geq 0 \text{ for } i = 1, 2\}.$$

The outcome of bargaining is either one of these agreements or disagreement. We assume that the players care only about the agreement they reach and possibly the time at which they reach it, not about the path of the negotiations that precede agreement. In particular, a player does not suffer if she agrees to an offer that is worse than one she previously rejected. Further, we assume that each

party regards a failure to reach an agreement as equivalent to obtaining none of the pie.

We study several models of bargaining. They differ in the specification of the order of moves and the options available to each player whenever she moves. As we will see, the details of the bargaining procedure critically affect the outcome of bargaining.

16.4.1 Take it or leave it (ultimatum game)

Player 1 proposes a division of the pie (a member of X), which player 2 then either accepts or rejects. (Example 16.1 is a version of this game in which the set of possible agreements is finite.)

Definition 16.9: Ultimatum game

The *ultimatum game* is the extensive game $\langle \{1,2\}, H, P, (\succsim^i)_{i\in\{1,2\}} \rangle$ with the following components.

Histories

The set H of histories consists of

- $\varnothing$ (the initial history)
- (x) for any $x \in X$ (player 1 makes the proposal x)
- (x, Y) for any $x \in X$ (player 1 makes the proposal x, which player 2 accepts)
- (x, N) for any $x \in X$ (player 1 makes the proposal x, which player 2 rejects).

Player function

$P(\varnothing) = 1$ (player 1 moves at the start of the game) and $P(x) = 2$ for all $x \in X$ (player 2 moves after player 1 makes a proposal).

Preferences

The preference relation $\succsim^i$ of each player i is represented by the payoff function u^i with $u^i(x, Y) = x^i$ and $u^i(x, N) = 0$ for all $x \in X$.

The game is illustrated in Figure 16.5. Note that this diagram (unlike the diagrams of previous games) does not show all the histories. It represents player 1's set of (infinitely many) actions by a shaded triangle, and shows only one of her actions, x, and the actions available to player 2 after the history (x).

Player 1's set of strategies in the ultimatum game is X, and each strategy of player 2 is a function that for each $x \in X$ specifies either Y or N. The game has

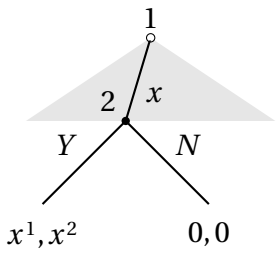


Figure 16.5 An illustration of the ultimatum game.

a unique subgame perfect equilibrium, in which player 1 proposes that she gets the entire pie and player 2 accepts all proposals.

Proposition 16.4: SPE of ultimatum game

The ultimatum game has a unique subgame perfect equilibrium, in which player 1 proposes $(1,0)$ and player 2 accepts all proposals.

Proof

The strategy pair is a subgame perfect equilibrium: given that player 2 accepts all proposals, the proposal $(1,0)$ is optimal for player 1, and after any proposal, acceptance (Y) is optimal for player 2.

Now let s be a subgame perfect equilibrium. The only optimal response of player 2 to a proposal x with $x^2 > 0$ is acceptance, so $s^2(x) = Y$ for any x with $x^2 > 0$. Thus if $s^1(\varnothing) = (1-\varepsilon, \varepsilon)$ with $\varepsilon > 0$ (which player 2 accepts), then player 1 can do better by proposing $(1 - \frac{1}{2}\varepsilon, \frac{1}{2}\varepsilon)$, which player 2 also accepts. Hence $s^1(\varnothing) = (1,0)$. Finally, $s^2(1,0) = Y$ because if player 2 rejects $(1,0)$ then player 1 gets 0 and can do better by making any other proposal, which player 2 accepts.

Notice that after the proposal $(1,0)$, player 2 is indifferent between Y and N. Nevertheless, the game has only one subgame perfect equilibrium, in which player 2 accepts $(1,0)$. In contrast, in Example 16.1, where the number of possible agreements is finite, the game has also a subgame perfect equilibrium in which player 2 rejects the proposal $(1,0)$.

A long history of experiments has demonstrated that the unique subgame perfect equilibrium is inconsistent with human behavior. For example, a population of 19,000 students from around the world similar to the readership of this book have responded to a question on `http://gametheory.tau.ac.il` asking them to imagine they have to divide \$100 between themselves and another person. The most common proposal, chosen by about 50% of subjects, is the

division (\$50, \$50). Only about 11% choose to offer the other person \$0 or \$1, as in the subgame perfect equilibrium.

Rather than necessarily calling into question the concept of subgame perfect equilibrium, these results point to the unrealistic nature of the players' preferences in the game. First, some people have preferences for fairness, which lead them to most prefer an equal division of the pie. Second, many people are insulted by low offers, and hence reject them. When players' preferences involve such considerations, the (modified) game may have a subgame perfect equilibrium in which the proposer receives significantly less than the entire pie.

16.4.2 Finite horizon with alternating offers

Now assume that after player 2 rejects player 1's offer, she can make a counteroffer, which player 1 can accept or reject, and that the players can continue to alternate proposals in this way for up to T periods. If the offer made in period T is rejected, the game ends with disagreement.

Definition 16.10: Finite-horizon bargaining game with alternating offers

A *finite-horizon bargaining game with alternating offers* is an extensive game $\langle \{1,2\}, H, P, (\succsim^i)_{i\in\{1,2\}} \rangle$ with the following components.

Histories

The set H of histories consists, for some positive integer T, of

- $\varnothing$ (the initial history)
- $(x_1, N, x_2, N, \ldots, x_t)$ for any $x_1, \ldots, x_t \in X$ and $1 \le t \le T$ (proposals through period $t-1$ are rejected, and the proposal in period t is x_t)
- $(x_1, N, x_2, N, \ldots, x_t, N)$ for any $x_1, \ldots, x_t \in X$ and $1 \le t \le T$ (proposals through period t are rejected)
- $(x_1, N, x_2, N, \ldots, x_{t-1}, N, x_t, Y)$ for any $x_1, \ldots, x_t \in X$ and $1 \le t \le T$ (proposals through period $t-1$ are rejected, and the proposal in period t is accepted).

Player function

Let $i_\tau = 1$ if τ is odd and $i_\tau = 2$ if τ is even. Then

- $P(\varnothing) = 1$ (player 1 makes the first proposal)
- $P(x_1, N, x_2, N, \ldots, x_t) = i_{t+1}$ for $t = 1, \ldots, T$ (player i_{t+1} responds to the offer made by i_t)

- $P(x_1, N, x_2, N, \ldots, x_t, N) = i_{t+1}$ for $t = 1, \ldots, T-1$ (player i_{t+1} makes the proposal at the beginning of period $t+1$).

Preferences

The preference relation $\succsim^i$ of each player i is represented by the payoff function u^i with $u^i(x_1, N, x_2, N, \ldots, x_t, Y) = x_t^i$ and $u^i(x_1, N, x_2, N, \ldots, x_T, N) = 0$.

We show that in this game all the bargaining power belongs to the player who makes the proposal in the last period: in every subgame perfect equilibrium this player receives the whole pie.

Proposition 16.5: SPE of finite-horizon game with alternating offers

In every subgame perfect equilibrium of a finite-horizon bargaining game with alternating offers, the payoff of the player who makes a proposal in the last period is 1 and the payoff of the other player is 0.

Proof

Let i be the player who proposes in period T, let j be the other player, and let $e(i)$ be the partition in which i gets the whole pie. The game has a subgame perfect equilibrium in which player i proposes $e(i)$ whenever she makes a proposal and accepts only $e(i)$ whenever she responds to a proposal, and player j always proposes $e(i)$ and accepts all proposals.

The game has many subgame perfect equilibria but all of them end with i getting all the pie. Let s be a subgame perfect equilibrium. Consider a history $h = (x_1, N, x_2, N, \ldots, x_{T-1}, N)$ in which $T-1$ proposals are made and rejected. The argument in the proof of Proposition 16.4 implies that $s^i(h) = e(i)$ and that player j accepts any proposal of player i in period T.

Now if i does not get the whole pie in the outcome of s, she can deviate profitably to the strategy r^i in which she rejects any proposal in any period and always proposes $e(i)$. The outcome of the pair of strategies r^i and s^j is agreement on $e(i)$ in period T at the latest.

16.4.3 Infinite horizon with one-sided offers

The result in the previous section demonstrates the significance of the existence of a final period in which a proposal can be made. We now study a model in

which no such final period exists: the players believe that after any rejection there will be another opportunity to agree. For now, we assume that only player 1 makes proposals. Since we do not limit the number of bargaining periods, we need to use a natural extension of the model of an extensive game in which terminal histories can be infinite.

Definition 16.11: Infinite-horizon bargaining game with one-sided offers

The *infinite-horizon bargaining game with one-sided offers* is the extensive game $\langle \{1,2\}, H, P, (\succsim^i)_{i\in\{1,2\}} \rangle$ with the following components.

Histories

The set H of histories consists of

- $\varnothing$ (the initial history)
- $(x_1, N, x_2, N, \ldots, x_{t-1}, N, x_t)$ for any $x_1, \ldots, x_t \in X$ and $t \geq 1$ (proposals through period $t-1$ are rejected, and player 1 proposes x_t in period t)
- $(x_1, N, x_2, N, \ldots, x_t, N)$ for any $x_1, \ldots, x_t \in X$ and $t \geq 1$ (proposals through period t are rejected)
- $(x_1, N, x_2, N, \ldots, x_{t-1}, N, x_t, Y)$ for any $x_1, \ldots, x_t \in X$ and $t \geq 1$ (proposals through period $t-1$ are rejected, and player 1's proposal in period t is accepted).
- $(x_1, N, x_2, N, \ldots, x_t, N, \ldots)$ for any infinite sequence of proposals $x_1, \ldots, x_t, \ldots$ (all proposals are rejected).

Player function

$P(\varnothing) = P(x_1, N, x_2, N, \ldots, x_t, N) = 1$ and $P(x_1, N, x_2, N, \ldots, x_t) = 2$.

Preferences

The preference relation $\succsim^i$ of each player i is represented by the payoff function u^i with

$$u^i(x_1, N, x_2, N, \ldots, x_t, Y) = x_t^i \text{ and } u^i(x_1, N, x_2, N, \ldots, x_t, N, \ldots) = 0.$$

In this game, every partition of the pie is the outcome of some subgame perfect equilibrium. In fact, for every partition of the pie, the game has a subgame perfect equilibrium in which agreement is reached immediately on that partition. Thus when the horizon is infinite, the fact that only player 1 makes offers does not give her more bargaining power than player 2.

Proposition 16.6: SPE of infinite-horizon game with one-sided offers

For every partition $x_* \in X$, the infinite-horizon bargaining game with one-sided offers has a subgame perfect equilibrium in which the outcome is immediate agreement on x_*. The game has also a subgame perfect equilibrium in which the players never reach agreement.

Proof

We first show that the following strategy pair is a subgame perfect equilibrium in which the players reach agreement in period 1 on x_*.

Player 1
Always propose x_*.

Player 2
Accept an offer y if and only if $y^2 \geq x_*^2$.

After the initial history or any history ending with rejection, player 1 can do no better than follow her strategy, because player 2 never accepts less than x_*^2. After any history ending with an offer y for which $y^2 < x_*^2$, player 2 can do no better than follow her strategy, because if she rejects the proposal then player 1 subsequently continue to propose x_*. After any history ending with an offer y for which $y^2 \geq x_*^2$, player 2 can do no better than follow her strategy and accept the proposal, because player 1 never proposes that player 2 gets more than x_*^2. Thus the strategy pair is a subgame perfect equilibrium.

We now show that the following strategy pair is a subgame perfect equilibrium in which the players never reach agreement.

Player 1
After the initial history and any history in which all proposals are $(1,0)$, propose $(1,0)$. After any other history, propose $(0,1)$.

Player 2
After any history in which all proposals are $(1,0)$, reject the proposal. After any other history, accept the proposal only if it is $(0,1)$.

Consider first player 1. After each history, if player 1 follows her strategy the outcome is either agreement on $(0,1)$ or disagreement, both of which yield the payoff 0. Any change in player 1's strategy after any history also generates disagreement or agreement on $(0,1)$, and thus does not make her better off.

Now consider player 2. After a history in which player 1 has proposed only $(1,0)$, player 2's following her strategy leads to the players' never reaching agreement, and any change in her strategy leads either to the same outcome or to agreement on $(1,0)$, which is no better for her. After a history in which player 1 has proposed a partition different from $(1,0)$, player 2's following her strategy leads her favorite agreement, $(0,1)$.

The equilibrium in which the players never reach agreement may be interpreted as follows. Initially, player 2 expects player 1 to insist on getting the whole pie and she plans to reject such a proposal. When player 1 makes any other proposal, player 2 interprets the move as a sign of weakness on the part of player 1 and expects player 1 to yield and offer her the whole pie. This interpretation of an attempt by player 1 to reach an agreement by offering player 2 a positive amount of the pie deters player 1 from doing so.

16.4.4 Infinite horizon with one-sided offers and discounting

We now modify the model in the previous section by assuming that each player prefers to receive pie earlier than later. Specifically, we assume that the payoff of each player i at a terminal history in which agreement on x is reached at time t is $(\delta^i)^t x^i$, where $\delta^i \in (0,1)$.

Definition 16.12: Infinite-horizon bargaining game with one-sided offers and discounting

An *infinite-horizon bargaining game with one-sided offers and discounting* is an extensive game that differs from an infinite-horizon bargaining game with one-sided offers only in that the payoff of player i to an agreement on x in period t is $(\delta^i)^t x^i$ for $i = 1, 2$, where $\delta^i \in (0,1)$. For notational economy we write $\delta^1 = \alpha$ and $\delta^2 = \beta$.

The first strategy pair in the proof of Proposition 16.6, in which player 1 always proposes x_* and player 2 accepts only proposals in which she receives at least x_*^2, is not a subgame perfect equilibrium of this game unless $x_*^2 = 0$. If $x_*^2 > 0$, consider the history in which at the beginning of the game player 1 proposes $(x_*^1 + \varepsilon, x_*^2 - \varepsilon)$ with $\varepsilon > 0$ small enough that $x_*^2 - \varepsilon > \beta x_*^2$. Given player 1's strategy, player 2's strategy (which in particular rejects the proposal) gives her x_*^2 at a later period; accepting $x_*^2 - \varepsilon$ is better for player 2.

We now show that the introduction of discounting makes a huge difference to the set of subgame perfect equilibria: it restores the bargaining power of the player who makes all offers, even if player 2's discount factor is close to 1.

Proposition 16.7: SPE of infinite-horizon game with one-sided offers and discounting

For any values of the discount factors α and β, an infinite-horizon bargaining game with one-sided offers and discounting has a unique subgame perfect equilibrium, in which player 1 gets all the pie immediately.

Proof

First note that the strategy pair in which player 1 always proposes $(1,0)$ and player 2 accepts all proposals is a subgame perfect equilibrium.

Now let M be the supremum of player 2's payoffs over all subgame perfect equilibria. Consider a history (x) (player 1 proposes x). If player 2 rejects x, the remainder of the game is identical to the whole game. Thus in any subgame perfect equilibrium any strategy that rejects x yields player 2 at most M with one period of delay. Hence in a subgame perfect equilibrium player 2 accepts x if $x^2 > \beta M$. So the infimum of player 1's payoffs over all subgame perfect equilibria is at least $1-\beta M$. Therefore the supremum of player 2's payoffs M does not exceed βM, which is possible only if $M=0$ (given $\beta<1$).

Given that player 2's payoff in every subgame perfect equilibrium is 0, she accepts all offers x in which $x^2>0$. Player 1's payoff in every subgame perfect equilibrium is 1 because for any strategy pair in which her payoff is $u<1$ she can deviate and propose $(u+\varepsilon, 1-u-\varepsilon)$ with $\varepsilon<1-u$, which player 2 accepts. Thus in any subgame perfect equilibrium player 1 offers $(1,0)$ and player 2 accepts all offers.

16.4.5 *Infinite horizon with alternating offers and discounting*

Finally consider a model in which the horizon is infinite, the players alternate offers, and payoffs obtained after period 1 are discounted.

Definition 16.13: Infinite-horizon bargaining game with alternating offers and discounting

An *infinite-horizon bargaining game with alternating offers and discounting* is an extensive game $\langle \{1,2\}, H, P, (\succsim^i)_{i\in\{1,2\}}\rangle$ with the following components, where $i_t=1$ if t is odd and $i_t=2$ if t is even.

Histories

The set H of histories consists of

- $\varnothing$ (the initial history)
- $(x_1, N, x_2, N, \ldots, x_t)$ for any $x_1, \ldots, x_t \in X$ and $t \geq 1$ (proposals through period $t-1$ are rejected, and the proposal in period t is x_t)
- $(x_1, N, x_2, N, \ldots, x_t, N)$ for any $x_1, \ldots, x_t \in X$ and $t \geq 1$ (proposals through period t are rejected)
- $(x_1, N, x_2, N, \ldots, x_{t-1}, N, x_t, Y)$ for any $x_1, \ldots, x_t \in X$ and $t \geq 1$ (proposals through period $t-1$ are rejected and the proposal in period t is accepted)
- $(x_1, N, x_2, N, \ldots, x_t, N, \ldots)$ for any infinite sequence of proposals $x_1, \ldots, x_t, \ldots$ (all proposals are rejected).

Player function

The player function is defined as follows

- $P(\varnothing) = 1$ (player 1 makes the first proposal)
- $P(x_1, N, x_2, N, \ldots, x_t) = i_{t+1}$ for $t \geq 1$ (player i_{t+1} responds to the offer made by i_t)
- $P(x_1, N, x_2, N, \ldots, x_t, N) = i_{t+1}$ for $t \geq 1$ (player i_{t+1} makes the proposal at the beginning of period $t+1$).

Preferences

The preference relation $\succsim^i$ of each player i is represented by the payoff function u^i with $u^i(x_1, N, x_2, N, \ldots, x_t, Y) = (\delta^i)^t x_t^i$ for $t \geq 1$ and $u^i(x_1, N, x_2, N, \ldots, x_t, N, \ldots) = 0$ for $i = 1, 2$, where $\delta^i \in (0,1)$. For notational economy we write $\delta^1 = \alpha$ and $\delta^2 = \beta$.

Giving player 2 the opportunity to make offers restores her bargaining power. We now show that the game has a unique subgame perfect equilibrium, in which the players' payoffs depend on their discount factors.

Proposition 16.8: SPE of infinite-horizon game with alternating offers and discounting

An infinite-horizon bargaining game with alternating offers and discounting has a unique subgame perfect equilibrium, in which

- player 1 always proposes x_* and accepts a proposal x if and only if $x^1 \geq y_*^1$

- player 2 always proposes y_* and accepts a proposal y if and only if $y^2 \geq x_*^2$

where

$$x_* = \left(\frac{1-\beta}{1-\alpha\beta}, \frac{\beta(1-\alpha)}{1-\alpha\beta}\right) \quad \text{and} \quad y_* = \left(\frac{\alpha(1-\beta)}{1-\alpha\beta}, \frac{1-\alpha}{1-\alpha\beta}\right).$$

Proof

Note that the pair of proposals x_* and y_* is the unique solution of the pair of equations $\alpha x^1 = y^1$ and $\beta y^2 = x^2$.

Step 1 *The strategy pair is a subgame perfect equilibrium.*

Proof. First consider a history after which player 1 makes a proposal. If player 1 follows her strategy, she proposes x_*, which player 2 accepts, resulting in player 1's getting x_*^1 immediately. Given player 2's strategy, player 1 can, by changing her strategy, either obtain an agreement not better than x_* in a later period or induce perpetual disagreement. Thus she has no profitable deviation.

Now consider a history after which player 1 responds to a proposal y.

If $y^1 \geq y_*^1$, player 1's strategy calls for her to accept the proposal, resulting in her getting y^1 immediately. If she deviates (and in particular rejects the proposal), then the outcome is not better for her than getting x_* at least one period later. Thus any deviation generates for her a payoff of at most $\alpha x_*^1 = y_*^1 \leq y^1$, so that she is not better off deviating from her strategy.

If $y_1 < y_*^1$, player 1's strategy calls for her to reject the proposal, in which case she proposes x_*, which player 2 accepts, resulting in x_* one period later. Any deviation leads her to either accept the proposal or to obtain offers not better than x_* at least one period later. Given $\alpha x_*^1 = y_*^1$, she is thus not better off accepting the proposal.

The argument for player 2 is similar. ◁

Step 2 *No other strategy pair is a subgame perfect equilibrium.*

Proof. Let G^i be the game following a history after which player i makes a proposal. (All such games are identical.) Let M^i be the supremum of player i's payoffs in subgame perfect equilibria of G^i and let m^i be the infimum of these payoffs.

We first argue that $m^2 \geq 1 - \alpha M^1$. If player 1 rejects player 2's initial proposal in G^2, play continues to G^1, in which player 1's payoff is at most M^1. Thus in any subgame perfect equilibrium player 1 optimally accepts any proposal that gives her more than αM^1, so that player 2's payoff in any equilibrium of G^2 is not less than $1 - \alpha M^1$. Hence $m^2 \geq 1 - \alpha M^1$.

We now argue that $M^1 \leq 1 - \beta m^2$. If player 2 rejects player 1's initial proposal in G^1, play continues to G^2, in which player 2's payoff is at least m^2. Thus player 2 optimally rejects any proposal that gives her less than βm^2, so that in no subgame perfect equilibrium of G^1 is player 1's payoff higher than $1 - \beta m^2$. Hence $M^1 \leq 1 - \beta m^2$.

These two inequalities imply that $1 - \alpha M^1 \leq m^2 \leq (1 - M^1)/\beta$ and hence $M^1 \leq (1-\beta)/(1-\alpha\beta) = x_*^1$. By Step 1, $M^1 \geq x_*^1$. Thus $M^1 = x_*^1$. Similar arguments yield $m^2 = y_*^2$.

Repeating these arguments with the roles of players 1 and 2 reversed yields $M^2 = y_*^2$ and $m^1 = x_*^1$, so that in every subgame perfect equilibrium of G^1 the payoff of player 1 is x_*^1 and in every subgame perfect equilibrium of G^2 the payoff of player 2 is y_*^2.

Now, in G^1 player 2, by rejecting player 1's proposal, can get at least $\beta y_*^2 = x_*^2$. Thus in every subgame perfect equilibrium of G^1 her payoff is x_*^2. Payoffs of x_*^1 for player 1 and x_*^2 for player 2 are possible only if agreement is reached immediately on x_*, so that in every subgame perfect equilibrium of G^1 player 1 proposes x_* and player 2 accepts this proposal. Similarly, in every subgame perfect equilibrium of G^2 player 2 proposes y^* and player 1 accepts this proposal. Thus the strategy pair given in the proposition is the only subgame perfect equilibrium of the game. $\triangleleft$

Notice that as a player values future payoffs more (becomes more patient), given the discount factor of the other player, the share of the pie that she receives increases. As her discount factor approaches 1, her equilibrium share approaches 1, regardless of the other player 2's (given) discount factor.

If the players are equally patient, with $\alpha = \beta = \delta$, the equilibrium payoff of player 1 is $1/(1+\delta)$ and that of player 2 is $\delta/(1+\delta)$. Thus the fact that player 1 makes the first proposal confers on her an advantage, but one that diminishes as both players become more patient. When δ is close to 1, the equilibrium payoff of each player is close to $\frac{1}{2}$. That is, when the players are equally patient and value future payoffs almost as much as they value current payoffs, the unique subgame perfect equilibrium involves an almost equal split of the pie.

In the subgame perfect equilibrium, agreement is reached immediately. In Problem 11 you are asked to analyze the game with different preferences: the

payoff of each player i for an agreement on x in period t is $x^i - c^i t$ for some c^1, $c^2 > 0$ (rather than $(\delta^i)^t x^i$). When $c^1 \neq c^2$ this game also has a unique subgame perfect equilibrium in which agreement is reached immediately. However, as you are asked to show in Problem 12b, when $c^1 = c^2$ the game has subgame perfect equilibria in which agreement is reached after a delay.

Problems

1. *Trust game.* Player 1 starts with \$10. She has to decide how much to keep and how much to transfer to player 2. Player 2 triples the amount of money she gets from player 1 and then decides how much, from that total amount, to transfer to player 1. Assume that each player is interested only in the amount of money she has at the end of the process.

   Model the situation as an extensive game and find its subgame perfect equilibria.

   Does the game have a Nash equilibrium outcome that is not a subgame perfect equilibrium outcome?

   What are the subgame perfect equilibria of the game in which the process is repeated three times?

2. *Multiple subgame perfect equilibria.* Construct an extensive game with two players that has two subgame perfect equilibria, one better for both players than the other.

3. *Nash equilibrium and subgame perfect equilibrium.* Construct an extensive game with two players that has a unique subgame perfect equilibrium and a Nash equilibrium that both players prefer to the subgame perfect equilibrium.

4. *Comparative statics.* Construct two extensive games that differ only in the payoff of one player, say player 1, regarding one outcome, such that each game has a unique subgame perfect equilibrium and the subgame perfect equilibrium payoff of player 1 is lower in the game in which her payoff is higher.

5. *Auction.* Two potential buyers compete for an indivisible item worth \$12. Buyer 1 has \$9 and buyer 2 has \$6. The seller will not accept any offer less than \$3.

   The buyers take turns bidding, starting with buyer 1. All bids are whole dollars and cannot exceed \$12. A player can bid more than the amount of cash

she holds, but if she wins she is punished severely, an outcome worse for her than any other. When one of the bidders does not raise the bid, the auction is over and the other player gets the item for the amount of her last bid.

a. Show that in all subgame perfect equilibria of the extensive game that models this situation player 1 gets the item.

b. Show that the game has a subgame perfect equilibrium in which the item is sold for \$3.

c. Show that the game has a subgame perfect equilibrium in which the item is sold for \$8.

6. *Solomon's mechanism.* An object belongs to one of two people, each of whom claims ownership. The value of the object is H to the owner and L to the other individual, where $H > L > 0$. King Solomon orders the two people to play the following game. Randomly, the people are assigned to be player 1 and player 2. Player 1 starts and has to declare either *mine* or *hers*. If she says *hers*, the game is over and player 2 gets the object. If she says *mine*, player 2 has to say either *hers*, in which case the object is given to player 1, or *mine*, in which case player 2 gets the object and pays M to King Solomon, with $H > M > L$, and player 1 pays a small amount $\varepsilon > 0$ to King Solomon.

   Explain why the outcome of this procedure is that the owner gets the object without paying anything.

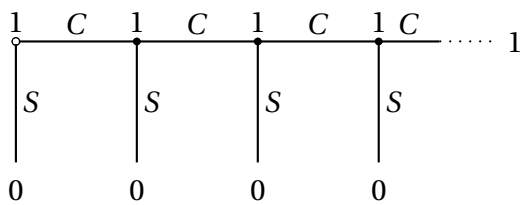
7. *Communication.* Consider a group of n people, with $n \geq 3$, living in separate locations, who need to share information that is received initially only by player 1. Assume that the information is beneficial only if all the players receive it. (The group may be a number of related families, the information may be instructions on how to get to a family gathering, and the gathering may be a success only if everyone attends.) When a player receives the information, she is informed of the path the information took. She then decides whether to pass the information to one of the players who has not yet received it. If every player receives the information, then every player who passed it on receives a payoff of $1 - c$, where $c \in (0,1)$, and the single player who got it last receives a payoff of 1. Otherwise, every player receives a payoff of $-c$ if she passed on the information and 0 otherwise.

   a. Draw the game tree for the case of $n = 3$.

   b. Characterize the subgame perfect equilibria of the game for each value of n.

8. *Race*. Two players, 1 and 2, start at distances A and B steps from a target. The player who reaches the target first gets a prize of P (and the other player gets no prize). The players alternate turns, starting with player 1. On her turn, a player can stay where she is, at a cost of 0, advance one step, at a cost of 2, or advance two steps, at a cost of 4.5. If for two successive turns both players stay where they are then the game ends (and neither player receives a prize). Each player aims to maximize her net gain (prize, if any, minus cost). For any values of A, B, and P with $6 \leq P \leq 8$, find the unique subgame perfect equilibrium of the extensive game that models this situation.

9. *One-deviation property*. Show that a strategy pair that satisfies the one-deviation property is not necessarily a subgame perfect equilibrium of a game that does not have a finite horizon by contemplating the one-player game illustrated below. In each period 1, 2, ... the player can *stop* or *continue*. If in any period she chooses *stop* then the game ends and her payoff is 0, whereas if she chooses *continue* she has another opportunity to *stop*. If she never chooses *stop* then her payoff is 1.



10. *Implementation*. You are a mediator in a case in which two neighbors cannot agree how to split the \$100 cost of hiring a gardener for their common property. Denote by v^i the value of hiring the gardener for neighbor i. Each neighbor knows both v^1 and v^2.

    Your aim is to ensure that the gardener is hired if the sum of the neighbors' values is greater than \$100 and is not hired if the sum is less than \$100.

    You suggest that the neighbors participate in the following procedure. First neighbor 1 names an amount. Neighbor 2 observes this amount and names an amount herself. (Each amount can be any nonnegative number.) If the sum of the amounts is less than \$100, the gardener is not hired. If the sum exceeds \$100, the gardener is hired and each neighbor pays the gardener the amount she named.

    a. Show that for any values of v^1 and v^2, the outcome of the subgame perfect equilibrium of the game that models this procedure is that the gardener is hired if $v^1 + v^2 > 100$ and is not hired if $v^1 + v^2 < 100$.

b. Show that if the neighbors are asked to report amounts simultaneously and the gardener is hired if any only if the sum of the reports is at least 100, then if $100 < v^1 + v^2 < 200$ and $v^i \leq 100$ for $i = 1, 2$ then the strategic game that models the procedure has a Nash equilibrium in which the gardener is not hired.

11. *Alternating offers with fixed bargaining costs.* Analyze the variant of the infinite-horizon bargaining game with alternating offers and discounting in which each player i values an amount x received in period t by $x - c^i t$ (rather than $(\delta^i)^t x$). (That is, she bears a fixed cost of c^i for each period that passes before agreement is reached.) Assume that $0 < c^1 < c^2$. Show that the game has a unique subgame perfect equilibrium and in this equilibrium player 1 gets the entire pie.

12. *Alternating offers with equal fixed costs.* Consider the variant of the game in Problem 11 in which $c^1 = c^2 = c$.

    a. Show that each partition in which player 1 gets at least c is an outcome of a subgame perfect equilibrium.

    b. Show that if $c \leq \frac{1}{3}$ then the game has a subgame perfect equilibrium in which agreement is not reached in period 1.

Notes

The notion of an extensive game originated with von Neumann and Morgenstern (1944); Kuhn (1950, 1953) suggested the model we describe. The notion of subgame perfect equilibrium is due to Selten (1965). Proposition 16.3 is due to Kuhn (1953).

The centipede game (Example 16.2) is due to Rosenthal (1981). Proposition 16.8 is due to Rubinstein (1982).

The game in Problem 6 is taken from Glazer and Ma (1989). The game in Problem 8 is a simplification due to Vijay Krishna of the model in Harris and Vickers (1985).