AN INTRODUCTION TO GAME THEORY

MARTIN J. OSBORNE
University of Toronto

NEW YORK OXFORD
OXFORD UNIVERSITY PRESS
2004

ters 1 and 6). Poundstone (1992, 257–272) writes informally about the game and its possible applications. The result in Exercise 177.2 is due to Becker (1974); see also Bergstrom (1989). The first formal study of chess is Zermelo (1913); see Schwalbe and Walker (2000) for a discussion of this paper and related work. Exercises 179.1, 179.2, and 179.3 are taken from Gardner (1959, Chapter 4), which includes several other intriguing examples.

---

6 Extensive Games with Perfect Information: Illustrations

Prerequisite: Chapter 5

THE first three sections of this chapter illustrate the notion of subgame perfect equilibrium in games in which the longest history has length two or three. The last section studies a game with an arbitrary finite horizon. Some games with infinite horizons are studied in Chapters 14, 15, and 16.

6.1 The ultimatum game, the holdup game, and agenda control

6.1.1 The ultimatum game

Bargaining over the division of a pie may naturally be modeled as an extensive game. Here I analyze a very simple game that is the basis of a richer model studied in Chapter 16. The game is so simple, in fact, that you may not initially think of it as a model of "bargaining".

Two people use the following procedure to split \$$c$. Person 1 offers person 2 an amount of money up to \$$c$. If 2 accepts this offer, then 1 receives the remainder of the \$$c$. If 2 rejects the offer, then *neither* person receives any payoff. Each person cares *only* about the amount of money she receives, and (naturally!) prefers to receive as much as possible.

Assume that the amount person 1 offers can be any number, not necessarily an integral number of cents. Then the following extensive game, known as the **ultimatum game**, models the procedure.

Players The two people.

Terminal histories The set of sequences (x, Z), where x is a number with $0 \leq x \leq c$ (the amount of money that person 1 offers to person 2) and Z is either Y ("yes, I accept") or N ("no, I reject").

Player function $P(\varnothing) = 1$ and $P(x) = 2$ for all x.

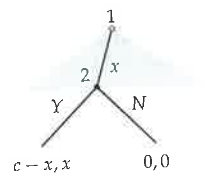


Figure 182.1 An illustration of the ultimatum game. The gray triangle represents the continuum of possible offers of player 1; the black lines indicate the terminal histories that start with the offer x.

Preferences Each person's preferences are represented by payoffs equal to the amounts of money she receives. For the terminal history (x, Y) person 1 receives $c - x$ and person 2 receives x; for the terminal history (x, N) each person receives 0.

This game is illustrated in Figure 182.1, in which the continuum of offers of player 1 is represented by the gray triangle, and the black lines indicate the terminal histories that start with the offer x. The game has a finite horizon, so we can use backward induction to find its subgame perfect equilibria. First consider the subgames of length 1, in which person 2 either accepts or rejects an offer of person 1. For every possible offer of person 1, there is such a subgame. In the subgame that follows an offer x of person 1 for which $x > 0$, person 2's optimal action is to accept (if she rejects, she gets nothing). In the subgame that follows the offer $x = 0$, person 2 is indifferent between accepting and rejecting. Thus in a subgame perfect equilibrium person 2's strategy either accepts all offers (including 0), or accepts all offers $x > 0$ and rejects the offer $x = 0$.

Now consider the whole game. For each possible subgame perfect equilibrium strategy of person 2, we need to find the optimal strategy of person 1.

- If person 2 accepts all offers (including 0), then person 1's optimal offer is 0 (which yields her the payoff \$$c$).
- If person 2 accepts all offers except zero, then *no* offer of person 1 is optimal! No offer $x > 0$ is optimal, because the offer $x/2$ (for example) is better, given that person 2 accept both offers. And an offer of 0 is not optimal because person 2 rejects it, leading to a payoff of 0 for person 1, who is thus better off offering any positive amount less than \$$c$.

We conclude that the only subgame perfect equilibrium of the game is the strategy pair in which person 1 offers 0 and person 2 accepts all offers. In this equilibrium, person 1's payoff is \$$c$ and person 2's payoff is zero.

This one-sided outcome is a consequence of the one-sided structure of the game. If we allow person 2 to make a counteroffer after rejecting person 1's opening offer (and possibly allow further responses by both players), so that the model corresponds more closely to a "bargaining" situation, then under some circumstances the outcome is less one-sided. (An extension of this type is explored in Chapter 16.)

EXERCISE 183.1 (Nash equilibria of the ultimatum game) Find the values of x for which there is a Nash equilibrium of the ultimatum game in which person 1 offers x.

EXERCISE 183.2 (Subgame perfect equilibria of the ultimatum game with indivisible units) Find the subgame perfect equilibria of the variant of the ultimatum game in which the amount of money is available only in multiples of a cent.

EXERCISE 183.3 (Dictator game and impunity game) The "dictator game" differs from the ultimatum game only in that person 2 does not have the option to reject person 1's offer (and thus has no strategic role in the game). The "impunity game" differs from the ultimatum game only in that person 1's payoff when person 2 rejects any offer x is $c - x$, rather than 0. (The game is named for the fact that person 2 is unable to "punish" person 1 for making a low offer.) Find the subgame perfect equilibria of each game.

EXERCISE 183.4 (Variants of ultimatum game and impunity game with equity-conscious players) Consider variants of the ultimatum game and impunity game in which each person cares not only about the amount of money she receives, but also about the equity of the allocation. Specifically, suppose that person i's preferences are represented by the payoff function given by $u_i(x_1, x_2) = x_i - \beta_i|x_1 - x_2|$, where x_i is the amount of money person i receives, $\beta_i > 0$, and, for any number z, $|z|$ denotes the absolute value of z (i.e. $|z| = z$ if $z > 0$ and $|z| = -z$ if $z < 0$). Assume $c = 1$. Find the set of subgame perfect equilibria of each game and compare them. Are there any values of β_1 and β_2 for which an offer is rejected in equilibrium? (An interesting further variant of the ultimatum game in which person 1 is uncertain about the value of β_2 is considered in Exercise 227.1.)

EXPERIMENTS ON THE ULTIMATUM GAME

The sharp prediction of the notion of subgame perfect equilibrium in the ultimatum game lends itself to experimental testing. The first test was conducted in the late 1970s among graduate students of economics in a class at the University of Cologne (in what was then West Germany). The amount c available varied among the games played; it ranged from 4 DM to 10 DM (around U.S.\$2 to U.S.\$5 at the time). A group of 42 students was split into two groups and seated on different sides of a room. Each member of one subgroup played the role of player 1 in an ultimatum game. She wrote down on a form the amount (up to c) that she demanded. Her form was then given to a randomly determined member of the other group, who, playing the role of player 2, either accepted what remained of the amount c or rejected it (in which case neither player received any payoff). Each player had 10 minutes to make her decision. The entire experiment was repeated a week later. (Güth, Schmittberger, and Schwarze 1982.)

In the first experiment the average demand by people playing the role of player 1 was 0.65c, and in the second experiment it was 0.69c, much less than the amount c or $c - 0.01$ predicted by the notion of subgame perfect equilibrium (0.01 DM was the smallest monetary unit; see Exercise 183.2). Almost 20% of offers were rejected over the two experiments, including one of 3 DM (out of a pie of 7 DM) and five of around 1 DM (out of pies of between 4 DM and 6 DM). Many other experiments, including one in which the amount of money to be divided was much larger (Hoffman, McCabe, and Smith 1996), have produced similar results. In brief, the results do not accord well with the predictions of subgame perfect equilibrium.

Or do they? Each player in the ultimatum game cares only about the amount of money she receives. But an experimental subject may care also about the amount of money her opponent receives. Further, a variant of the ultimatum game in which the players are equity conscious has subgame perfect equilibria in which offers are significant (as you will have discovered if you did Exercise 183.4).

However, if people are equity conscious in the strategic environment of the ultimatum game, they are presumably equity conscious also in related environments; an explanation of the experimental results in the ultimatum game based on the players' preferences' exhibiting equity conscience is not convincing if it applies only to that environment. Several related games have been studied, among them the dictator game and the impunity game (Exercise 183.3). In the subgame perfect equilibria of these games, player 1 offers 0; in a variant in which the players are equity conscious, player 1's offers are no higher than they are in the analogous variant of the ultimatum game, and, for moderate degrees of equity conscience, are lower (see Exercise 183.4). These features of the equilibria are broadly consistent with the experimental evidence on dictator, impunity, and ultimatum games (see, for example, Forsythe, Horowitz, Savin, and Sefton 1994, Bolton and Zwick 1995, and Güth and Huck 1997).

One feature of the experimental results is inconsistent with subgame perfect equilibrium even when players are equity conscious (at least given the form of the payoff functions in Exercise 183.4): positive offers are sometimes rejected. The equilibrium strategy of an equity-conscious player 2 in the ultimatum game rejects inequitable offers, but, knowing this, player 1 does not, in equilibrium, make such an offer. To generate rejections in equilibrium we need to further modify the model by assuming that people differ in their degree of equity conscience, and that player 1 does not know the degree of equity conscience of player 2 (see Exercise 227.1).

An alternative explanation of the experimental results focuses on player 2's behavior. The evidence is consistent with player 1's significant offers in the ultimatum game being driven by a fear that player 2 will reject small offers—a fear that is rational, because small offers are often rejected. Why does player 2 behave in this way? One argument is that in our daily lives, we use "rules of thumb" that work well in the situations in which we are typically involved; we do not calculate our rational actions in each situation. Further, we are not typically involved in one-shot situations with the structure of the ultimatum game. Instead, we usually engage in repeated interactions, where it is advantageous to "punish" a player who makes a paltry offer, and to build a reputation for not accepting such offers. Experimental subjects may apply such rules of thumb rather than carefully thinking through the logic of the game, and thus reject low offers in an ultimatum game but accept them in an impunity game, where rejection does not affect the proposer. The experimental evidence so far collected is broadly consistent with both this explanation and the explanation based on the nature of players' preferences.

EXERCISE 185.1 (Bargaining over two indivisible objects) Consider a variant of the ultimatum game, with indivisible units. Two people use the following procedure to allocate two desirable identical indivisible objects. One person proposes an allocation (both objects go to person 1, both go to person 2, one goes to each person), which the other person then either accepts or rejects. In the event of rejection, neither person receives either object. Each person cares only about the number of objects she obtains. Construct an extensive game that models this situation and find its subgame perfect equilibria. Does the game have any Nash equilibrium that is not a subgame perfect equilibrium? Is there any outcome that is generated by a Nash equilibrium but not by any subgame perfect equilibrium?

EXERCISE 185.2 (Dividing a cake fairly) Two players use the following procedure to divide a cake. Player 1 divides the cake into two pieces, and then player 2 chooses one of the pieces; player 1 obtains the remaining piece. The cake is continuously divisible (no lumps!), and each player likes all parts of it.

a. Suppose that the cake is perfectly homogeneous, so that each player cares only about the size of the piece of cake she obtains. How is the cake divided in a subgame perfect equilibrium?

b. Suppose that the cake is not homogeneous: the players evaluate different parts of it differently. Represent the cake by the set C, so that a piece of the cake is a subset P of C. Assume that if P is a subset of P' not equal to P' (smaller than P'), then each player prefers P' to P. Assume also that the players' preferences are continuous: if player i prefers P to P', then there is a subset of P not equal to P that player i also prefers to P'. Let (P_1, P_2) (where P_1 and P_2 together constitute the whole cake C) be the division chosen by player 1 in a subgame perfect equilibrium of the divide-and-choose game, P_2 being the piece chosen by player 2. Show that player 2 is indifferent between P_1 and P_2, and player 1 likes P_1 at least as much as P_2. Give an example in which player 1 prefers P_1 to P_2.

6.1.2 *The holdup game*

Before engaging in an ultimatum game in which she may accept or reject an offer of person 1, person 2 takes an action that affects the size c of the pie to be divided. She may exert little effort, resulting in a small pie, of size c_L, or great effort, resulting in a large pie, of size c_H. She dislikes exerting effort. Specifically, assume that her payoff is $x - E$ if her share of the pie is x, where $E = L$ if she exerts little effort and $E = H > L$ if she exerts great effort. The extensive game that models this situation is known as the **holdup game**.

EXERCISE 186.1 (Holdup game) Formulate the holdup game precisely. (Write down the set of players, the set of terminal histories, the player function, and the players' preferences.)

What is the subgame perfect equilibrium of the holdup game? Each subgame that follows person 2's choice of effort is an ultimatum game, and thus has a unique subgame perfect equilibrium, in which person 1 offers 0 and person 2 accepts all offers. Now consider person 2's choice of effort at the start of the game. If she chooses L, then her payoff, given the outcome in the following subgame, is $-L$, whereas if she chooses H, then her payoff is $-H$. Consequently she chooses L. Thus the game has a unique subgame perfect equilibrium, in which person 2 exerts little effort and person 1 obtains all of the resulting small pie.

This equilibrium does not depend on the values of c_L, c_H, L, and H (given that $H > L$). In particular, even if c_H is much larger than c_L, but H is only slightly larger than L, person 2 exerts little effort in the equilibrium, although both players could be much better off if person 2 were to exert great effort (which, in this case, is not very great) and person 2 were to obtain some of the extra pie. No such superior outcome is sustainable in an equilibrium because person 2, having exerted great effort, may be "held up" for the entire pie by person 1.

This result does not depend sensitively on the extreme subgame perfect equilibrium outcome of the ultimatum game. A similar result emerges when the bargaining following person 2's choice of effort generates a more equal division of the pie. By exerting great effort, player 2 increases the size of the pie. The point is that if the negotiation results in some (not necessarily all) of this extra pie going to player 1, then for some values of player 2's cost of exerting great effort less than the value of the extra pie, player 2 prefers to exert little effort. In these circumstances, player 2's exerting great effort generates outcomes in which both players are better off than they are when player 2 exerts little effort, but because the bargaining puts some of the extra pie in the hands of player 1, player 2's incentive is to exert little effort.

6.1.3 *Agenda control*

In some legislatures, proposals for modifications of the law are formulated by committees. Under a "closed rule", the legislature may either accept or reject a proposed modification, but may not propose an alternative; in the event of rejection, the existing law is unchanged. That is, the committee controls the "agenda". (In Section 10.9 I consider a reason why a legislature might cede such power to a committee.)

Model an outcome as a number y. Assume that the legislature and committee have favorite outcomes that may differ, and that the preferences of each body are represented by a single-peaked payoff function symmetric about its favorite outcome, like the voters' preferences in Hotelling's model of electoral competition (see Figure 71.1). Assign numbers to outcomes so that the legislature's favorite outcome is 0; denote the committee's favorite outcome by $y_c > 0$. Then the following variant of the ultimatum game models the procedure. The players are the committee and the legislature. The committee proposes an outcome y, which the legislature either accepts or rejects. In the event of rejection the outcome is y_0, the "status quo". Note that the main respect in which this game differs from the ultimatum game is that the players' preferences are diametrically opposed only with regard to outcomes between 0 and y_c; if $y' < y'' < 0$ or $y_c < y'' < y'$, then both players prefer y'' to y'.

EXERCISE 187.1 (Agenda control) Find the subgame perfect equilibrium of this game as a function of the status quo outcome y_0. Show, in particular, that for a range of values of y_0, an increase in the value of y_0 leads to a *decrease* in the value of the equilibrium outcome.

6.2 Stackelberg's model of duopoly

6.2.1 *General model*

In the models of oligopoly in Sections 3.1 and 3.2, each firm chooses its action not knowing the other firms' actions. How do the conclusions change when the firms move sequentially? Is a firm better off moving before or after the other firms?

In this section I consider a market in which there are two firms, both producing the same good. Firm i's cost of producing q_i units of the good is $C_i(q_i)$; the price at which output is sold when the total output is Q is $P_d(Q)$. (In Section 3.1 I denote this function P; here I add a d subscript to avoid a conflict with the player function of the extensive game.) Each firm's strategic variable is output, as in Cournot's model (Section 3.1), but the firms make their decisions sequentially, rather than simultaneously: one firm chooses its output, then the other firm does so, knowing the output chosen by the first firm.

We can model this situation by the following extensive game, known as **Stackelberg's duopoly game** (after an early analyst of duopoly with asynchronous actions).

Players The two firms.

Terminal histories The set of all sequences (q_1, q_2) of outputs for the firms (where each q_i, the output of firm i, is a nonnegative number).

Player function $P(\varnothing) = 1$ and $P(q_1) = 2$ for all q_1.

Preferences The payoff of firm i to the terminal history (q_1, q_2) is its profit $q_iP_d(q_1 + q_2) - C_i(q_i)$, for $i = 1, 2$.

Firm 1 moves at the start of the game. Thus a strategy of firm 1 is simply an output. Firm 2 moves after every history in which firm 1 chooses an output. Thus a strategy of firm 2 is a *function* that associates an output for firm 2 with each possible output of firm 1.

The game has a finite horizon, so we may use backward induction to find its subgame perfect equilibria.

- First, for any output of firm 1, we find the outputs of firm 2 that maximize its profit. Suppose that for each output q_1 of firm 1 there is one such output of firm 2; denote it $b_2(q_1)$. Then in any subgame perfect equilibrium, firm 2's strategy is b_2.

- Next, we find the outputs of firm 1 that maximize its profit, *given the strategy of firm 2*. When firm 1 chooses the output q_1, firm 2 chooses the output $b_2(q_1)$, resulting in a total output of $q_1 + b_2(q_1)$, and hence a price of $P_d(q_1 + b_2(q_1))$. Thus firm 1's output in a subgame perfect equilibrium is a value of q_1 that maximizes
$$q_1P_d(q_1 + b_2(q_1)) - C_1(q_1). \tag{188.1}$$
Suppose that there is one such value of q_1; denote it q_1^*.

We conclude that if firm 2 has a unique best response $b_2(q_1)$ to each output q_1 of firm 1, and firm 1 has a unique best action q_1^*, given firm 2's best responses, then the subgame perfect equilibrium of the game is (q_1^*, b_2): firm 1's equilibrium strategy is q_1^* and firm 2's equilibrium strategy is the function b_2. The output chosen by firm 2, given firm 1's equilibrium strategy, is $b_2(q_1^*)$; denote this output q_2^*.

When firm 1 chooses any output q_1, the outcome, given that firm 2 uses its equilibrium strategy, is the pair of outputs $(q_1, b_2(q_1))$. That is, as firm 1 varies its output, the outcome varies along firm 2's best response function b_2. Thus we can characterize the subgame perfect equilibrium outcome (q_1^*, q_2^*) as the point on firm 2's best response function that maximizes firm 1's profit.

6.2.2 *Example: constant unit cost and linear inverse demand*

Suppose that $C_i(q_i) = cq_i$ for $i = 1, 2$, and
$$P_d(Q) = \begin{cases} \alpha - Q & \text{if } Q \leq \alpha \\ 0 & \text{if } Q > \alpha, \end{cases} \tag{188.2}$$
where $c > 0$ and $c < \alpha$ (as in the example of Cournot's duopoly game in Section 3.1.3). We found that under these assumptions firm 2 has a unique best response to each output q_1 of firm 1, given by
$$b_2(q_1) = \begin{cases} \frac{1}{2}(\alpha - c - q_1) & \text{if } q_1 \leq \alpha - c \\ 0 & \text{if } q_1 > \alpha - c. \end{cases}$$

Thus in a subgame perfect equilibrium of Stackelberg's game firm 2's strategy is this function b_2 and firm 1's strategy is the output q_1 that maximizes
$$q_1(\alpha - c - (q_1 + \tfrac{1}{2}(\alpha - c - q_1))) = \tfrac{1}{2}q_1(\alpha - c - q_1)$$
(refer to (188.1)). This function is a quadratic in q_1 that is zero when $q_1 = 0$ and when $q_1 = \alpha - c$. Thus its maximizer is $q_1 = \frac{1}{2}(\alpha - c)$.

We conclude that the game has a unique subgame perfect equilibrium, in which firm 1's strategy is the output $\frac{1}{2}(\alpha - c)$ and firm 2's strategy is b_2. The outcome of the equilibrium is that firm 1 produces the output $q_1^* = \frac{1}{2}(\alpha - c)$ and firm 2 produces the output $q_2^* = b_2(q_1^*) = b_2(\frac{1}{2}(\alpha - c)) = \frac{1}{2}(\alpha - c - \frac{1}{2}(\alpha - c)) = \frac{1}{4}(\alpha - c)$. Firm 1's profit is $q_1^*(P_d(q_1^* + q_2^*) - c) = \frac{1}{8}(\alpha - c)^2$, and firm 2's profit is $q_2^*(P_d(q_1^* + q_2^*) - c) = \frac{1}{16}(\alpha - c)^2$. By contrast, in the unique Nash equilibrium of Cournot's (simultaneous-move) game under the same assumptions, each firm produces $\frac{1}{3}(\alpha - c)$ units of output and obtains the profit $\frac{1}{9}(\alpha - c)^2$. Thus under our assumptions firm 1 produces more output and obtains more profit in the subgame perfect equilibrium of the sequential game in which it moves first than it does in the Nash equilibrium of Cournot's game, and firm 2 produces less output and obtains less profit.

(?) EXERCISE 189.1 (Stackelberg's duopoly game with quadratic costs) Find the subgame perfect equilibrium of Stackelberg's duopoly game when $C_i(q_i) = q_i^2$ for $i = 1, 2$, and $P_d(Q) = \alpha - Q$ for all $Q \leq \alpha$ (with $P_d(Q) = 0$ for $Q > \alpha$). Compare the equilibrium outcome with the Nash equilibrium of Cournot's game under the same assumptions (Exercise 59.1).

6.2.3 *Properties of subgame perfect equilibrium*

First-mover's equilibrium profit In the example just studied, the first-mover is better off in the subgame perfect equilibrium of Stackelberg's game than it is in the Nash equilibrium of Cournot's game. A weak version of this result holds under very general conditions: for any cost and inverse demand functions for which firm 2 has a unique best response to each output of firm 1, firm 1 is at least as well off in any subgame perfect equilibrium of Stackelberg's game as it is in any Nash equilibrium of Cournot's game. This result follows from the general result in Exercise 177.3*a*. The argument is simple. One of firm 1's options in Stackelberg's game is to choose its output in some Nash equilibrium of Cournot's game. If it chooses such an output, then firm 2's best action is to choose its output in the same Nash equilibrium, given the assumption that it has a unique best response to each output of firm 1. Thus by choosing such an output, firm 1 obtains its profit at a Nash equilibrium of Cournot's game; by choosing a different output it may possibly obtain a higher payoff.

Equilibrium outputs In the example in the previous section (6.2.2), firm 1 produces more output in the subgame perfect equilibrium of Stackelberg's game than it does

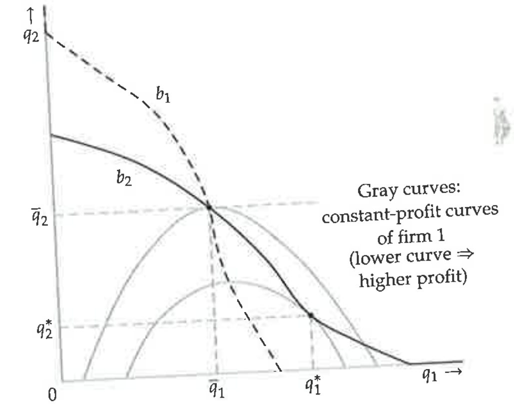


Figure 190.1 The subgame perfect equilibrium outcome (q_1^*, q_2^*) of Stackelberg's game and the Nash equilibrium $(\bar{q}_1, \bar{q}_2)$ of Cournot's game. Along each gray curve, firm 1's profit is constant; the lower curve corresponds to higher profit than does the upper curve. Each curve has a slope of zero where it crosses firm 1's best response function b_1.

in the Nash equilibrium of Cournot's game, and firm 2 produces less. A weak form of this result holds whenever firm 2's best response function is decreasing where it is positive (i.e. a higher output for firm 1 implies a lower optimal output for firm 2).

The argument is illustrated in Figure 190.1. The firms' best response functions are the curves labeled b_1 (dashed) and b_2. The Nash equilibrium of Cournot's game is the intersection $(\bar{q}_1, \bar{q}_2)$ of these curves. Along each gray curve, firm 1's profit is constant; the *lower* curve corresponds to a *higher* profit. (For any given value of firm 1's output, a reduction in the output of firm 2 increases the price and thus increases firm 1's profit.) Each constant-profit curve of firm 1 is horizontal where it crosses firm 1's best response function, because the best response is precisely the output that maximizes firm 1's profit, given firm 2's output. (Cf. Figure 61.1.) Thus the subgame perfect equilibrium outcome—the point on firm 2's best response function that yields the highest profit for firm 1—is the point (q_1^*, q_2^*) in Figure 190.1. In particular, given that the best response function of firm 2 is downward sloping, firm 1 produces at least as much, and firm 2 produces at most as much, in the subgame perfect equilibrium of Stackelberg's game as in the Nash equilibrium of Cournot's game.

For some cost and demand functions, firm 2's output in a subgame perfect equilibrium of Stackelberg's game is zero. An example is shown in Figure 191.1. The discontinuity in firm 2's best response function at q_1^* in this example may arise because firm 2 incurs a "fixed" cost—a cost independent of its output—when it produces a positive output (see Exercise 59.2). When firm 1's output is q_1^*, firm 2's

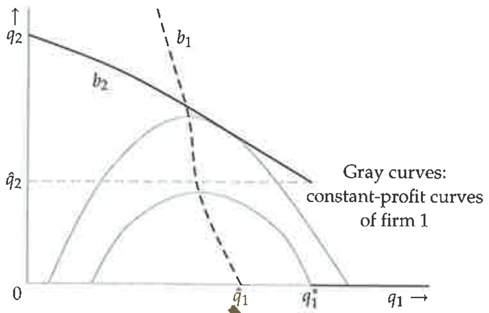


Figure 191.1 The subgame perfect equilibrium output q_1^* of firm 1 in Stackelberg's sequential game when firm 2 incurs a fixed cost. Along each gray curve, firm 1's profit is constant; the lower curve corresponds to higher profit than does the upper curve.

maximal profit is zero, which it obtains both when it produces no output (and does not pay the fixed cost) and when it produces the output $\hat{q}_2$. When firm 1 produces less than q_1^*, firm 2's maximal profit is positive, and firm 2 optimally produces a positive output; when firm 1 produces more than q_1^*, firm 2 optimally produces no output. Given this form of firm 2's best response function and the form of firm 1's constant-profit curves in Figure 190.1, the point on firm 2's best response function that yields firm 1 the highest profit is $(q_1^*, 0)$.

I claim that this example has a unique subgame perfect equilibrium, in which firm 1 produces q_1^* and firm 2's strategy coincides with its best response function except at q_1^*, where the strategy specifies the output 0. The output firm 2's equilibrium strategy specifies after each history must be a best response to firm 1's output, so the only question regarding firm 2's strategy is whether it specifies an output of 0 or $\hat{q}_2$ when firm 1's output is q_1^*. The argument that there is no subgame perfect equilibrium in which firm 2's strategy specifies the output $\hat{q}_2$ is similar to the argument that there is no subgame perfect equilibrium in the ultimatum game in which person 2 rejects the offer 0. If firm 2 produces the output $\hat{q}_2$ in response to firm 1's output q_1^*, then firm 1 has no optimal output: it would like to produce a little more than q_1^*, inducing firm 2 to produce zero, but is better off the closer its output is to q_1^*. Because there is no smallest output greater than q_1^*, no output is *optimal* for firm 1 in this case. Thus the game has no subgame perfect equilibrium in which firm 2's strategy specifies the output $\hat{q}_2$ in response to firm 1's output q_1^*.

Note that if firm 2 were entirely absent from the market, firm 1 would produce $\hat{q}_1$, less than q_1^*. Thus firm 2's presence affects the outcome, even though it produces no output.

⊛ EXERCISE 191.1 (Stackelberg's duopoly game with fixed costs) Suppose that the inverse demand function is given by (188.2) and the cost function of each firm i is given by

$$C_i(q_i) = \begin{cases} 0 & \text{if } q_i = 0 \\ f + cq_i & \text{if } q_i > 0, \end{cases}$$

where $c \geq 0$, $f > 0$, and $c < \alpha$, as in Exercise 59.2. Show that if $c = 0$, $\alpha = 12$, and $f = 4$, Stackelberg's game has a unique subgame perfect equilibrium, in which firm 1's output is 8 and firm 2's output is zero. (Use your results from Exercise 59.2).

The value of commitment Firm 1's output in a subgame perfect equilibrium of Stackelberg's game is *not* in general a best response to firm 2's output: if firm 1 could adjust its output after firm 2 has chosen its output, then it would do so! (In the case shown in Figure 190.1, it would reduce its output.) However, if firm 1 had this opportunity, and firm 2 knew that it had the opportunity, then firm 2 would choose a different output. Indeed, if we simply add a third stage to the game, in which firm 1 chooses an output, then the first stage is irrelevant, and *firm 2* is effectively the first-mover; in the subgame perfect equilibrium firm 1 is worse off than it is in the Nash equilibrium of the simultaneous-move game. (In the example in Section 6.2.2, the unique subgame perfect equilibrium has firm 2 choose the output $(\alpha - c)/2$ and firm 1 choose the output $(\alpha - c)/4$.) In summary, even though firm 1 can increase its profit by changing its output after firm 2 has chosen its output, in the game in which it has this opportunity it is worse off than it is in the game in which it must choose its output before firm 2 and cannot subsequently modify this output. That is, firm 1 prefers to be *committed* not to change its mind.

EXERCISE 192.1 (Sequential variant of Bertrand's duopoly game) Consider the variant of Bertrand's duopoly game (Section 3.2) in which first firm 1 chooses a price, then firm 2 chooses a price. Assume that each firm is restricted to choose a price that is an integral number of cents (as in Exercise 67.2), that each firm's unit cost is constant and equal to c (an integral number of cents), and that the monopoly profit is positive.

a. Specify an extensive game with perfect information that models this situation.

b. Give an example of a strategy of firm 1 and an example of a strategy of firm 2.

c. Find the subgame perfect equilibria of the game.

6.3 Buying votes

A legislature has k members, where k is an odd number. Two rival bills, X and Y, are being considered. The bill that attracts the votes of a majority of legislators will pass. Interest group X favors bill X, whereas interest group Y favors bill Y. Each group wishes to entice a majority of legislators to vote for its favorite bill. First interest group X gives an amount of money (possibly zero) to each legislator, then interest group Y does so. Each interest group wishes to spend as little as possible. Group X values the passing of bill X at $\$V_X > 0$ and the passing of bill Y at zero, and group Y values the passing of bill Y at $\$V_Y > 0$ and the passing of bill X at zero. (For example, group X is indifferent between an outcome in which it spends V_X and bill X is passed and one in which it spends nothing and bill Y is passed.) Each legislator votes for the favored bill of the interest group that offers her the most money; a legislator to whom both groups offer the same amount of money votes for bill Y (an arbitrary assumption that simplifies the analysis without qualitatively changing the outcome). For example, if $k = 3$, the amounts offered to the legislators by group X are $x = (100, 50, 0)$, and the amounts offered by group Y are $y = (100, 0, 50)$, then legislators 1 and 3 vote for Y and legislator 2 votes for X, so that Y passes. (In some actual legislatures the inducements offered to legislators are more subtle than cash transfers.)

We can model this situation as the following extensive game.

Players The two interest groups, X and Y.

Terminal histories The set of all sequences (x, y), where x is a list of payments to legislators made by interest group X and y is a list of payments to legislators made by interest group Y. (That is, both x and y are lists of k nonnegative integers.)

Player function $P(\varnothing) = X$ and $P(x) = Y$ for all x.

Preferences The preferences of interest group X are represented by the payoff function

$$\begin{cases} V_X - (x_1 + \cdots + x_k) & \text{if bill } X \text{ passes} \\ -(x_1 + \cdots + x_k) & \text{if bill } Y \text{ passes,} \end{cases}$$

where bill Y passes after the terminal history (x, y) if and only if the number of components of y that are at least equal to the corresponding components of x is at least $\frac{1}{2}(k+1)$ (a bare majority of the k legislators). The preferences of interest group Y are represented by the analogous function (where V_Y replaces V_X, y replaces x, and Y replaces X).

Before studying the subgame perfect equilibria of this game for arbitrary values of the parameters, consider two examples. First suppose that $k = 3$ and $V_X = V_Y = 300$. Under these assumptions, the most group X is willing to pay to get bill X passed is 300. For any payments it makes to the three legislators that sum to at most 300, two of the payments sum to at most 200, so that if group Y matches these payments it spends less than V_Y (= 300) and gets bill Y passed. Thus in any subgame perfect equilibrium group X makes no payments, group Y makes no payments, and (given the tie-breaking rule) bill Y is passed.

Now suppose that $k = 3$, $V_X = 300$, and $V_Y = 100$. In this case by paying each legislator more than 50, group X makes matching payments by group Y unprofitable: only by spending more than V_Y (= 100) can group Y cause bill Y to be passed. However, there is no subgame perfect equilibrium in which group X pays each legislator more than 50 because it can always pay a little less (as long as the payments still exceed 50) and still prevent group Y from profitably matching. In the only subgame perfect equilibrium group X pays each legislator exactly 50 and group Y makes no payments. Given group X's action, group Y is indifferent between matching X's payments (so that bill Y is passed) and making no payments. However, there is no subgame perfect equilibrium in which group Y matches group X's payments because if this were group Y's response, then group X could increase its payments a little, making matching payments by group Y unprofitable.

For arbitrary values of the parameters, the subgame perfect equilibrium outcome takes one of the forms in these two examples: either no payments are made and bill Y is passed, or group X makes payments that group Y does not wish to match, group Y makes no payments, and bill X is passed.

To find the subgame perfect equilibria in general, we may use backward induction. First consider group Y's best response to an arbitrary strategy x of group X. Let $\mu = \frac{1}{2}(k+1)$, a bare majority of k legislators, and denote by m_x the sum of the smallest μ components of x—the total payments Y needs to make to buy off a bare majority of legislators.

- If $m_x < V_Y$, then group Y can buy off a bare majority of legislators for less than V_Y, so that its best response to x is to match group X's payments to the μ legislators to whom group X's payments are smallest; the outcome is that bill Y is passed.

- If $m_x > V_Y$, then the cost to group Y of buying off any majority of legislators exceeds V_Y, so that group Y's best response to x is to make no payments; the outcome is that bill X is passed.

- If $m_x = V_Y$, then both the actions in the previous two cases are best responses by group Y to x.

We conclude that group Y's strategy in a subgame perfect equilibrium has the following properties.

- After a history x for which $m_x < V_Y$, group Y matches group X's payments to the μ legislators to whom X's payments are smallest.

- After a history x for which $m_x > V_Y$, group Y makes no payments.

- After a history x for which $m_x = V_Y$, group Y either makes no payments or matches group X's payments to the μ legislators to whom X's payments are smallest.

Given that group Y's subgame perfect equilibrium strategy has these properties, what should group X do? If it chooses a list of payments x for which $m_x < V_Y$, then group Y matches its payments to a bare majority of legislators, and bill Y passes. If it reduces all its payments, the same bill is passed. Thus the only list of payments x with $m_x < V_Y$ that may be optimal is $(0, \ldots, 0)$. If it chooses a list of payments x with $m_x > V_Y$, then group Y makes no payments, and bill X passes.

If it reduces all its payments a little (keeping the payments to every bare majority greater than V_Y), the outcome is the same. Thus no list of payments x for which $m_x > V_Y$ is optimal.

We conclude that in any subgame perfect equilibrium we have either $x = (0, \ldots, 0)$ (group X makes no payments) or $m_x = V_Y$ (the smallest sum of group X's payments to a bare majority of legislators is V_Y). Under what conditions does each case occur? If group X needs to spend more than V_X to deter group Y from matching its payments to a bare majority of legislators, then its best strategy is to make no payments ($x = (0, \ldots, 0)$). How much does it need to spend to deter group Y? It needs to pay more than V_Y to every bare majority of legislators, so it needs to pay each legislator more than V_Y/μ, in which case its total payment is more than kV_Y/μ. Thus if $V_X < kV_Y/\mu$, group X is better off making no payments than getting bill X passed by making payments large enough to deter group Y from matching its payments to a bare majority of legislators.

If $V_X > kV_Y/\mu$, on the other hand, group X can afford to make payments large enough to deter group Y from matching. In this case its best strategy is to pay each legislator V_Y/μ, so that its total payment to every bare majority of legislators is V_Y. Given this strategy, group Y is indifferent between matching group X's payments to a bare majority of legislators and making no payments. I claim that the game has no subgame perfect equilibrium in which group Y matches. The argument is similar to the argument that the ultimatum game has no subgame perfect equilibrium in which person 2 rejects the offer 0. Suppose that group Y matches. Then group X can increase its payoff by increasing its payments a little (keeping the total less than V_X), thereby deterring group Y from matching, and ensuring that bill X passes. Thus in any subgame perfect equilibrium group Y makes no payments in response to group X's strategy.

In conclusion, if $V_X \neq kV_Y/\mu$, then the game has a unique subgame perfect equilibrium, in which group Y's strategy is to

- match group X's payments to the μ legislators to whom X's payments are smallest after a history x for which $m_x < V_Y$, and

- make no payments after a history x for which $m_x \geq V_Y$

and group X's strategy depends on the relative sizes of V_X and V_Y:

- if $V_X < kV_Y/\mu$, then group X makes no payments;

- if $V_X > kV_Y/\mu$, then group X pays each legislator V_Y/μ.

If $V_X < kV_Y/\mu$, then the outcome is that neither group makes any payment, and bill Y is passed; if $V_X > kV_Y/\mu$, then the outcome is that group X pays each legislator V_Y/μ, group Y makes no payments, and bill X is passed. (If $V_X = kV_Y/\mu$, then the analysis is more complex.)

Three features of the subgame perfect equilibrium are significant. First, the outcome favors the second-mover in the game (group Y): only if $V_X > kV_Y/\mu$, which is close to $2V_Y$ when k is large, does group X manage to get bill X passed. Second, group Y never makes any payments! According to its equilibrium strategy it is prepared to make payments in response to certain strategies of group X, but given group X's *equilibrium* strategy, it spends not a cent. Third, if group X makes any payments (as it does in the equilibrium for $V_X > kV_Y/\mu$), then it makes a payment to *every* legislator. If there were no competing interest group but nonetheless each legislator would vote for bill X only if she were paid at least some amount, then group X would make payments to only a bare majority of legislators; if it were to act in this way in the presence of group Y, it would supply group Y with almost a majority of legislators who could be induced to vote for bill Y at no cost.

EXERCISE 196.1 (Three interest groups buying votes) Consider a variant of the model in which there are *three* bills, X, Y, and Z, and *three* interest groups, X, Y, and Z, who choose lists of payments sequentially. Ties are broken in favor of the group moving later. Assume that if each bill obtains the vote of one legislator, then bill X passes. Find the bill passed in any subgame perfect equilibrium when $k = 3$ and (*a*) $V_X = V_Y = V_Z = 300$, (*b*) $V_X = 300$, $V_Y = V_Z = 100$, and (*c*) $V_X = 300$, $V_Y = 202$, $V_Z = 100$. (You may assume that in each case a subgame perfect equilibrium exists; note that you are not asked to find the subgame perfect equilibria themselves.)

EXERCISE 196.2 (Interest groups buying votes under supermajority rule) Consider another variant of the model in which a supermajority is required to pass a bill. There are two bills, X and Y, and a "default outcome". A bill passes if and only if it receives at least $k^* > \frac{1}{2}(k+1)$ votes; if neither bill passes, the default outcome occurs. There are two interest groups. Both groups attach value 0 to the default outcome. Find the bill that is passed in any subgame perfect equilibrium when $k = 7$, $k^* = 5$, and (*a*) $V_X = V_Y = 700$ and (*b*) $V_X = 750$, $V_Y = 400$. In each case, would the legislators be better off or worse off if a simple majority of votes were required to pass a bill?

EXERCISE 196.3 (Sequential positioning by two political candidates) Consider the variant of Hotelling's model of electoral competition in Section 3.3 in which the n candidates choose their positions sequentially, rather than simultaneously. Model this situation as an extensive game. Find the subgame perfect equilibrium (equilibria?) when $n = 2$.

EXERCISE 196.4 (Sequential positioning by three political candidates) Consider a further variant of Hotelling's model of electoral competition in which the n candidates choose their positions sequentially and each candidate has the option of staying out of the race. Assume that each candidate prefers to stay out than to enter and lose, prefers to enter and tie with any number of candidates than to stay out, and prefers to tie with as few other candidates as possible. Model the situation as an extensive game and find the subgame perfect equilibrium outcomes when $n = 2$ (easy) and when $n = 3$ and the voters' favorite positions are distributed uniformly from 0 to 1 (i.e. the fraction of the voters' favorite positions less than x is x) (hard).

6.4 A race

6.4.1 *General model*

Firms compete with each other to develop new technologies; authors compete with each other to write books and film scripts about momentous current events; scientists compete with each other to make discoveries. In each case the winner enjoys a significant advantage over the losers, and each competitor can, at a cost, increase her pace of activity. How do the presence of competitors and size of the prize affect the pace of activity? How does the identity of the winner of the race depend on each competitor's initial distance from the finish line?

We can model a race as an extensive game with perfect information in which the players alternately choose how many "steps" to take. Here I study a simple example of such a game, with two players.

Player i is initially $k_i > 0$ steps from the finish line, for $i = 1, 2$. On each of her turns, a player can either not take any steps (at a cost of 0), or can take one step, at a cost of $c(1)$, or two steps, at a cost of $c(2)$. The first player to reach the finish line wins a prize, worth $v_i > 0$ to player i; the losing player's payoff is 0. To make the game finite, I assume that if, on successive turns, neither player takes any step, the game ends and neither player obtains the prize.

I denote the game in which player i moves first by $G_i(k_1, k_2)$. The game $G_1(k_1, k_2)$ is defined precisely as follows.

Players The two parties.

Terminal histories The set of sequences of the form $(x^1, y^1, x^2, y^2, \ldots, x^T)$ or $(x^1, y^1, x^2, y^2, \ldots, y^T)$ for some integer T, where each x^t (the number of steps taken by player 1 on her tth turn) and each y^t (the number of steps taken by player 2 on her tth turn) is 0, 1, or 2, there are never two successive 0's except possibly at the end of a sequence, and either $x^1 + \cdots + x^T = k_1$ and $y^1 + \cdots + y^T < k_2$ (player 1 reaches the finish line first), or $x^1 + \cdots + x^T < k_1$ and $y^1 + \cdots + y^T = k_2$ (player 2 reaches the finish line first).

Player function $P(\varnothing) = 1$, $P(x^1) = 2$ for all x^1, $P(x^1, y^1) = 1$ for all (x^1, y^1), $P(x^1, y^1, x^2) = 2$ for all (x^1, y^1, x^2), and so on.

Preferences For a terminal history in which player i loses, her payoff is the negative of the sum of the costs of all her moves; for a terminal history in which she wins it is v_i minus the sum of these costs.

6.4.2 *Subgame perfect equilibria of an example*

A simple example illustrates the features of the subgame perfect equilibria of this game. Suppose that both v_1 and v_2 are between 6 and 7 (their exact values do not affect the equilibria), the cost $c(1)$ of a single step is 1, and the cost $c(2)$ of two steps is 4. (Given that $c(2) > 2c(1)$, each player, in the absence of a competitor, would like to take one step at a time.)

The game has a finite horizon, so we may use backward induction to find its subgame perfect equilibria. Each of its subgames is either a game $G_i(m_1, m_2)$ with $i = 1$ or $i = 2$ and $0 < m_1 \leq k_1$ and $0 < m_2 \leq k_2$, or, if the last player to move before the subgame took no steps, a game that differs from $G_i(m_1, m_2)$ only in that it ends if player i initially takes no steps (i.e. the only terminal history starting with 0 consists only of 0).

First consider the very simplest game, $G_1(1,1)$, in which each player is initially one step from the finish line. If player 1 takes one step, she wins; if she does not move, then player 2 optimally takes one step (if she does not, the game ends) and wins. We conclude that the game has a unique subgame perfect equilibrium, in which player 1 initially takes one step and wins.

A similar argument applies to the game $G_1(1,2)$. If player 1 does not move, then player 2 has the option of taking one or two steps. If she takes one step, then play moves to a subgame identical $G_1(1,1)$, in which we have just concluded that player 1 wins. Thus player 2 takes two steps, and wins, if player 1 does not move at the start of $G_1(1,2)$. We conclude that the game has a unique subgame perfect equilibrium, in which player 1 initially takes one step and wins.

Now consider player 1's options in the game $G_1(2,1)$.

- Player 1 takes two steps: she wins, and obtains a payoff of at least $6 - 4 = 2$ (her valuation is more than 6, and the cost of two steps is 4).
- Player 1 takes one step: play moves to a subgame identical to $G_2(1,1)$; we know that in the equilibrium of this subgame player 2 initially takes one step and wins.
- Player 1 does not move: play moves to a subgame in which player 2 is the first-mover and is one step from the finish line, and, if player 2 does not move, the game ends. In an equilibrium of this subgame, player 2 takes one step and wins.

We conclude that the game $G_1(2,1)$ has a unique subgame perfect equilibrium, in which player 1 initially takes two steps and wins.

I have spelled out the details of the analysis of these cases to show how we use the result for the game $G_1(1,1)$ to find the equilibria of the games $G_1(1,2)$ and $G_1(2,1)$. In general, the equilibria of the games $G_i(k_1, k_2)$ for all values of k_1 and k_2 up to $\bar{k}$ tell us the consequences of player 1's taking one or two steps in the game $G_1(\bar{k}+1, \bar{k})$.

EXERCISE 198.1 (The race $G_1(2,2)$) Show that the game $G_1(2,2)$ has a unique subgame perfect equilibrium outcome, in which player 1 initially takes two steps, and wins.

So far we have concluded that in any game in which each player is initially at most two steps from the finish line, the first-mover takes enough steps to reach the finish line, and wins.

Now suppose that player 1 is at most two steps from the finish line, but player 2 is three steps away. Suppose that player 1 takes only *one* step (even if she is initially two steps from the finish line). Then if player 2 takes either one or two steps, play moves to a subgame in which player 1 (the first-mover) wins. Thus player 2 is better off not moving (and not incurring any cost), in which case player 1 takes one step on her next turn, and wins. (Player 1 prefers to move one step at a time than to move two steps initially because the former costs her 2 whereas the latter costs her 4.) We conclude that the outcome of a subgame perfect equilibrium in the game $G_1(2,3)$ is that player 1 takes one step on her first turn, then player 2 does not move, and then player 1 takes another step, and wins.

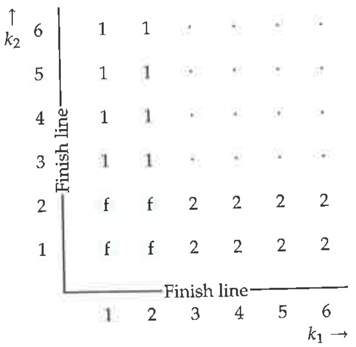
By a similar argument, in a subgame perfect equilibrium of any game in which player 1 is at most two steps from the finish line and player 2 is three or more steps away, player 1 moves one step at a time, and player 2 does not move; player 1 wins. Symmetrically, in a subgame perfect equilibrium of any game in which player 1 is three or more steps from the finish line and player 2 is at most two steps away, player 1 does not move, and player 2 moves one step at a time, and wins.

Our conclusions so far are illustrated in Figure 200.1, where player 1 moves to the left and player 2 moves down. The values of (k_1, k_2) for which the subgame perfect equilibrium outcome has been determined so far are labeled. The label "1" means that, regardless of who moves first, in a subgame perfect equilibrium player 1 moves one step on each turn, and player 2 does not move; player 1 wins. Similarly, the label "2" means that, regardless of who moves first, player 2 moves one step on each turn, and player 1 does not move; player 2 wins. The label "f" means that the first player to move takes enough steps to reach the finish line, and wins.

Now consider the game $G_1(3,3)$. If player 1 takes one step, we reach the game $G_2(2,3)$. From Figure 200.1 we see that in the subgame perfect equilibrium of this game player 1 wins, and does so by taking one step at a time (the point $(2,3)$ is labeled "1"). If player 1 takes two steps, we reach the game $G_2(1,3)$, in which player 1 also wins. Player 1 prefers not to take two steps unless she has to, so in the subgame perfect equilibrium of $G_1(3,3)$ she takes one step at a time, and wins, and player 2 does not move. Similarly, in a subgame perfect equilibrium of $G_2(3,3)$, player 2 takes one step at a time, and wins, and player 1 does not move.

A similar argument applies to each of the games $G_i(3,4)$, $G_i(4,3)$, and $G_i(4,4)$ for $i = 1, 2$. The argument differs only if the first-mover is four steps from the finish line, in which case she initially takes two steps to reach a game in which she wins. (If she initially takes only one step, the other player wins.)

Now consider the game $G_i(3,5)$ for $i = 1, 2$. By taking one step in $G_1(3,5)$, player 1 reaches a game in which she wins by taking one step at a time. The cost of her taking three steps is less than v_1, so in a subgame perfect equilibrium of $G_1(3,5)$



6	1	1	·	·	·	·
5	1	1	·	·	·	·
4	1	1	·	·	·	·
3	1	1	·	·	·	·
2	f	f	2	2	2	2
1	f	f	2	2	2	2
	1	2	3	4	5	6

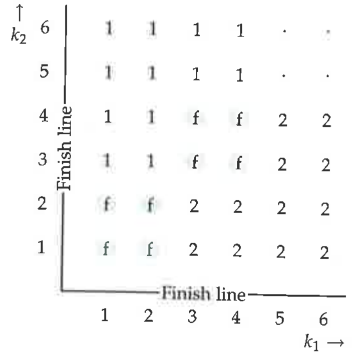
Figure 200.1 The subgame perfect equilibrium outcomes of the race $G_i(k_1, k_2)$. Player 1 moves to the left, and player 2 moves down. The values of (k_1, k_2) for which the subgame perfect equilibrium outcome has been determined so far are labeled; dots represent cases that have not yet been studied. The labels are explained in the text.

she takes one step at a time, and wins, and player 2 does not move. If player 2 takes either one or two steps in $G_2(3,5)$, she reaches a game (either $G_1(3,4)$ or $G_1(3,3)$) in which player 1 wins. Thus whatever she does, she loses, so that in a subgame perfect equilibrium she does not move and player 1 moves one step at a time. We conclude that in a subgame perfect equilibrium of both $G_1(3,5)$ and $G_2(3,5)$, player 1 takes one step on each turn and player 2 does not move; player 1 wins.

A similar argument applies to any game in which one player is initially three or four steps from the finish line and the other player is five or more steps from the finish line. We have now made arguments to justify the labeling in Figure 201.1, where the labels have the same meaning as in Figure 200.1, except that "f" means that the first player to move takes enough steps to reach the finish line *or* to reach the closest point labeled with her name, whichever is closer.

A feature of the subgame perfect equilibrium of the game $G_1(4,4)$ is noteworthy. Suppose that, as planned, player 1 takes two steps, but then player 2 deviates from her equilibrium strategy and takes two steps (rather than not moving). According to our analysis, player 1 should take two steps, to reach the finish line. If she does so, her payoff is negative (less than $7-4-4=-1$). Nevertheless she should definitely take the two steps: if she does not, her payoff is even smaller (-4), because player 2 wins. The point is that the cost of her first move is "sunk"; her decision after player 2 deviates must be based on her options from that point on.

The analysis of the games in which each player is initially either five or six steps from the finish line involves arguments similar to those used in the previous cases, with one amendment. A player who is initially six steps from the finish line is better off not moving at all (and obtaining the payoff 0) than she is moving two steps on any turn (and obtaining a negative payoff). An implication is that in the game $G_1(6,5)$, for example, player 1 does not move: if she takes only one step, then player 2 becomes the first-mover and, by taking a single step, moves the play to a game that she wins. We conclude that the first-mover wins in the games $G_i(5,5)$ and $G_i(6,6)$, whereas player 2 wins in $G_i(6,5)$ and player 1 wins in $G_i(5,6)$, for $i=1,2$.



6	1	1	1	1	·	·
5	1	1	1	1	·	·
4	1	1	f	f	2	2
3	1	1	f	f	2	2
2	f	f	2	2	2	2
1	f	f	2	2	2	2
	1	2	3	4	5	6

Figure 201.1 The subgame perfect equilibrium outcomes of the race $G_i(k_1, k_2)$. Player 1 moves to the left, and player 2 moves down. The values of (k_1, k_2) for which the subgame perfect equilibrium outcome has been determined so far are labeled; dots represent cases that have not yet been studied. The labels are explained in the text.

A player who is initially more than six steps from the finish line obtains a negative payoff if she moves, even if she wins, so in any subgame perfect equilibrium she does not move. Thus our analysis of the game is complete. The subgame perfect equilibrium outcomes are indicated in Figure 202.1, which shows also the steps taken in the equilibrium of each game when player 1 is the first-mover.

EXERCISE 201.1 (A race in which the players' valuations of the prize differ) Find the subgame perfect equilibrium outcome of the game in which player 1's valuation of the prize is between 6 and 7, and player 2's valuation is between 4 and 5.

In both of the following exercises, inductive arguments on the length of the game, like the one for $G_i(k_1, k_2)$, can be used.

EXERCISE 201.2 (Removing stones) Two people take turns removing stones from a pile of n stones. Each person may, on each of her turns, remove either one or two stones. The person who takes the last stone is the winner; she gets \$1 from her opponent. Find the subgame perfect equilibria of the games that model this situation for $n=1$ and $n=2$. Find the winner in each subgame perfect equilibrium for

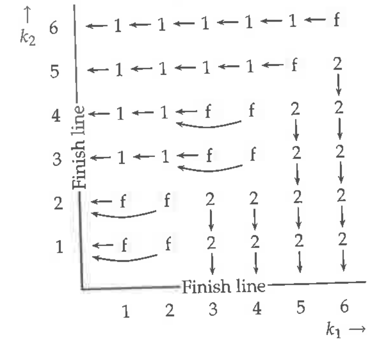


Figure 202.1 The subgame perfect equilibrium outcomes of the race $G_i(k_1, k_2)$. Player 1 moves to the left, and player 2 moves down. The arrows indicate the steps taken in the subgame perfect equilibrium outcome of the games in which player 1 moves first. The labels are explained in the text.

$n = 3$, using the fact that the subgame following player 1's removal of one stone is the game for $n = 2$ in which player 2 is the first-mover, and the subgame following player 1's removal of two stones is the game for $n = 1$ in which player 2 is the first-mover. Use the same technique to find the winner in each subgame perfect equilibrium for $n = 4$, and, if you can, for an arbitrary value of n.

EXERCISE 202.1 (Hungry lions) The members of a hierarchical group of hungry lions face a piece of prey. If lion 1 does not eat the prey, the prey escapes and the game ends. If it eats the prey, it becomes fat and slow, and lion 2 can eat it. If lion 2 does not eat lion 1, the game ends; if it eats lion 1, then it may be eaten by lion 3, and so on. Each lion prefers to eat than to be hungry, but prefers to be hungry than to be eaten. Find the subgame perfect equilibrium (equilibria?) of the extensive game that models this situation for any number n of lions.

6.4.3 *General lessons*

Each player's equilibrium strategy involves a "threat" to speed up if the other player deviates. Consider, for example, the game $G_1(3,3)$. Player 1's equilibrium strategy calls for her to take one step at a time, and player 2's equilibrium strategy calls for her not to move. Thus in the equilibrium outcome, player 1's debt climbs to 3 (the cost of her three single steps) before she reaches the finish line.

Now suppose that after player 1 takes her first step, player 2 deviates and takes a step. Then player 1's strategy calls for her to take two steps, raising her debt to 5. If at no stage can her debt exceed 3 (its maximal level if both players adhere to their equilibrium strategies), then her strategy cannot embody such threats.

The general point is that a limit on the debt a player can accumulate may affect the outcome even if it exceeds the player's debt in the equilibrium outcome in the absence of any limits. You are asked to study an example in the next exercise.

EXERCISE 203.1 (A race with a liquidity constraint) Find the subgame perfect equilibrium of the variant of the game $G_1(3,3)$ in which player 1's debt may never exceed 3.

In the subgame perfect equilibrium of every game $G_i(k_1, k_2)$, only one player moves; her opponent "gives up". This property of equilibrium holds in more general games. What added ingredient might lead to an equilibrium in which both players are active? A player's uncertainty about the other's characteristics would seem to be such an ingredient: if a player does not know the cost of its opponent's moves, it may assign a positive probability less than one to its winning, at least until it has accumulated some evidence of its opponent's behavior, and while it is optimistic it may be active even though its rival is also active. To build such considerations into the model we need to generalize the model of an extensive game to encompass imperfect information, as we do in Chapter 10.

Another robust feature of the subgame perfect equilibrium of $G_i(k_1, k_2)$ is that the presence of a competitor has little effect on the speed of the player who moves. A lone player would move one step at a time. When there are two players, for most starting points the one that moves does so at the same leisurely pace. Only for a small number of starting points, in all of which the players' initial distances from the starting line are similar, does the presence of a competitor induce the active player to hasten its progress, and then only in the first period.

Notes

The first experiment on the ultimatum game is reported in Güth, Schmittberger, and Schwarze (1982). Grout (1984) is an early analysis of a holdup game. The model of agenda control in legislatures is based on Denzau and Mackay (1983); Romer and Rosenthal (1978) earlier explored a similar idea. The model in Section 6.2 derives its name from the analysis in von Stackelberg (1934, Chapter 4). The vote-buying game in Section 6.3 is taken from Groseclose and Snyder (1996). The model of a race in Section 6.4 is a simplification suggested by Vijay Krishna of a model of Harris and Vickers (1985).

For more discussion of the experimental evidence on the ultimatum game (discussed in the box on page 183), see Roth (1995). Bolton and Ockenfels (2000) study the implications of assuming that players are equity conscious, and relate these implications to the experimental outcomes in various games. The explanation of the experimental results in terms of rules of thumb is discussed by Aumann (1997, 7–8). The problem of fair division, an example of which is given in Exercise 185.2, is studied in detail by Brams and Taylor (1996), who trace the idea of divide-and-choose back to antiquity (p. 10). I have been unable to find the origin of the idea in Exercise 202.1; Barton Lipman suggested the formulation in the exercise.

7 Extensive Games with Perfect Information: Extensions and Discussion

Prerequisite: Chapter 5

7.1 Allowing for simultaneous moves

7.1.1 Definition

THE model of an extensive game with perfect information (Definition 155.1) assumes that after every sequence of events, a single decision-maker takes an action, knowing every decision-maker's previous actions. I now describe a more general model that allows us to study situations in which, after some sequences of events, the members of a group of decision-makers choose their actions "simultaneously", each member knowing every decision-maker's *previous* actions, but not the contemporaneous actions of the other members of the group.

In the more general model, a terminal history is a sequence of *lists* of actions, each list specifying the actions of a set of players. (A game in which each set contains a single player is an extensive game with perfect information as defined previously.) For example, consider a situation in which player 1 chooses either C or D, then players 2 and 3 simultaneously take actions, each choosing either E or F. In the extensive game that models this situation, $(C,(E,E))$ is a terminal history in which first player 1 chooses C, and then players 2 and 3 both choose E. In the general model, the player function assigns a *set* of players to each nonterminal history. In the example just described, this set consists of the single player 1 for the empty history, and consists of players 2 and 3 for the history C.

An extensive game with perfect information (Definition 155.1) does not specify explicitly the sets of actions available to the players. However, we may derive the set of actions of the player who moves after any nonterminal history from the set of terminal histories and the player function (see (156.1)). When we allow simultaneous moves, the players' sets of actions are conveniently specified in the definition of a game. In the example of the previous paragraph, for instance, we specify the game by giving the eight possible terminal histories (C or D followed by one of the four pairs (E,E), (E,F), (F,E), and (F,F)), the player function defined by $P(\varnothing) = 1$ and $P(C) = P(D) = \{2,3\}$, the sets of actions $\{C,D\}$ for player 1 at the start of the game and $\{E,F\}$ for both player 2 and player 3 after the histories C and D, and each player's preferences over terminal histories.

In any game, the set of terminal histories, player function, and sets of actions for the players must be consistent: the list of actions that follows a subhistory of any terminal history must be a list of actions of the players assigned by the player function to that subhistory. In the game just described, for example, the list of actions following the subhistory C of the terminal history $(C,(E,E))$ is (E,E), which is a pair of actions for the players (2 and 3) assigned by the player function to the history C.

Precisely, an extensive game with perfect information and simultaneous moves is defined as follows.

▷ DEFINITION 206.1 An **extensive game with perfect information and simultaneous moves** consists of

- a set of **players**
- a set of sequences (**terminal histories**) with the property that no sequence is a proper subhistory of any other sequence
- a function (the **player function**) that assigns a set of players to every sequence that is a proper subhistory of some terminal history
- for each proper subhistory h of each terminal history and each player i that is a member of the set of players assigned to h by the player function, a set $A_i(h)$ (the set of **actions** available to player i after the history h)
- for each player, **preferences** over the set of terminal histories

such that the set of terminal histories, player function, and sets of actions are consistent in the sense that h is a terminal history if and only if either (*i*) h takes the form $(a^1,\ldots,a^k)$ for some integer k, the player function is not defined at h, and for every $\ell = 0,\ldots,k-1$, the element $a^{\ell+1}$ is a list of actions of the players assigned by the player function to $(a^1,\ldots,a^\ell)$ (the empty history if $\ell = 0$), or (*ii*) h takes the form $(a^1,a^2,\ldots)$ and for every $\ell = 0,1,\ldots$, the element $a^{\ell+1}$ is a list of actions of the players assigned by the player function to $(a^1,\ldots,a^\ell)$ (the empty history if $\ell = 0$).

This definition encompasses both extensive games with perfect information as in Definition 155.1 and, in a sense, strategic games. An extensive game with perfect information is an extensive game with perfect information and simultaneous moves in which the set of players assigned to each history consists of exactly one member. (The definition of an extensive game with perfect information and simultaneous moves includes the players' actions, whereas the definition of an extensive game with perfect information does not. However, actions may be derived from the terminal histories and player function of the latter.)

For any strategic game there is an extensive game with perfect information and simultaneous moves in which every terminal history has length 1 that models the same situation. In this extensive game, the set of terminal histories is the set of action profiles in the strategic game, the player function assigns the set of all players to the empty history, and the single set $A_i(\varnothing)$ of actions of each player i is the set of actions of player i in the strategic game.

◆ EXAMPLE 207.1 (Variant of *BoS*) First, person 1 decides whether to stay home and read a book or to attend a concert. If she reads a book, the game ends. If she decides to attend a concert, then, as in *BoS*, she and person 2 independently choose whether to sample the aural delights of Bach or Stravinsky, not knowing the other person's choice. Both people prefer to attend the concert of their favorite composer in the company of the other person to the outcome in which person 1 stays home and reads a book, and prefer this outcome to attending the concert of their less preferred composer in the company of the other person; the worst outcome for both people is that they attend different concerts.

The following extensive game with perfect information and simultaneous moves models this situation.

Players The two people (1 and 2).

Terminal histories *Book*, $(\textit{Concert},(B,B))$, $(\textit{Concert},(B,S))$, $(\textit{Concert},(S,B))$, and $(\textit{Concert},(S,S))$.

Player function $P(\varnothing) = 1$ and $P(\textit{Concert}) = \{1,2\}$.

Actions The set of player 1's actions at the empty history $\varnothing$ is $A_1(\varnothing) = \{\textit{Concert}, \textit{Book}\}$ and the set of her actions after the history *Concert* is $A_1(\textit{Concert}) = \{B,S\}$; the set of player 2's actions after the history *Concert* is $A_2(\textit{Concert}) = \{B,S\}$.

Preferences Player 1 prefers $(\textit{Concert},(B,B))$ to *Book* to $(\textit{Concert},(S,S))$ to $(\textit{Concert},(B,S))$, which she regards as indifferent to $(\textit{Concert},(S,B))$. Player 2 prefers $(\textit{Concert},(S,S))$ to *Book* to $(\textit{Concert},(B,B))$ to $(\textit{Concert},(B,S))$, which she regards as indifferent to $(\textit{Concert},(S,B))$.

This game is illustrated in Figure 208.1, in which I represent the simultaneous choices between B and S in the way that I previously represented a strategic game. (Only a game in which all the simultaneous moves occur at the end of terminal histories may be represented in a diagram like this one. For most other games no convenient diagrammatic representation exists.)

7.1.2 Strategies and Nash equilibrium

As in a game without simultaneous moves, a player's strategy specifies the action she chooses for every history after which it is her turn to move. Definition 159.1

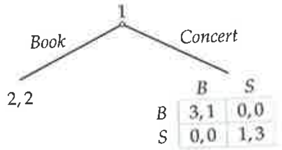


Figure 208.1 The variant of *BoS* described in Example 207.1.

requires only minor rewording to allow for the possibility that players may move simultaneously.

DEFINITION 208.1 (*Strategy in extensive game with perfect information and simultaneous moves*) A **strategy** of player i in an extensive game with perfect information and simultaneous moves is a function that assigns to each history h after which i is one of the players whose turn it is to move (i.e. i is a member of $P(h)$, where P is the player function of the game) an action in $A_i(h)$ (the set of actions available to player i after h).

The definition of a *Nash equilibrium* of an extensive game with perfect information and simultaneous moves is exactly the same as the definition for a game with no simultaneous moves (Definition 161.2): a Nash equilibrium is a strategy profile with the property that no player can induce a better outcome for herself by changing her strategy, given the other players' strategies. Also as before, the *strategic form* of a game is the strategic game in which the players' actions are their strategies in the extensive game (see Section 5.3), and a strategy profile is a Nash equilibrium of the extensive game if and only if it is a Nash equilibrium of the strategic form of the game.

EXAMPLE 208.2 (Nash equilibria of a variant of *BoS*) In the game in Example 207.1, a strategy of player 1 specifies her actions at the start of the game and after the history *Concert*; a strategy of player 2 specifies her action after the history *Concert*. Thus player 1 has four strategies, $(\textit{Concert}, B)$, $(\textit{Concert}, S)$, $(\textit{Book}, B)$, and $(\textit{Book}, S)$, and player 2 has two strategies, B and S. (Remember that a player's strategy is more than a plan of action; it specifies an action for *every* history after which the player moves, even histories that it precludes. For example, player 1's strategy specifies her action after the history *Concert* even if it specifies that she choose *Book* at the beginning of the game.)

The strategic form of the game is given in Figure 209.1. We see that the game has three pure Nash equilibria: $((\textit{Concert}, B), B)$, $((\textit{Book}, B), S)$, and $((\textit{Book}, S), S)$.

Every extensive game has a unique strategic form. However, some strategic games are the strategic forms of more than one extensive game. Consider, for example, the strategic game in Figure 209.2. This game is the strategic form of the extensive game with perfect information and simultaneous moves in which the two players choose their actions simultaneously; it is also the strategic form of the entry game in Figure 156.1.

	B	S
$(\textit{Concert}, B)$	3, 1	0, 0
$(\textit{Concert}, S)$	0, 0	1, 3
$(\textit{Book}, B)$	2, 2	2, 2
$(\textit{Book}, S)$	2, 2	2, 2

Figure 209.1 The strategic form of the game in Example 207.1.

	L	R
T	1, 2	1, 2
B	0, 0	2, 0

Figure 209.2 A strategic game that is the strategic form of more than one extensive game.

7.1.3 Subgame perfect equilibrium

As for a game in which one player moves after each history, the subgame following the history h of an extensive game with perfect information and simultaneous moves is the extensive game "starting at h". (The formal definition is a variant of Definition 164.1.)

For instance, the game in Example 207.1 has two subgames: the whole game, and the game in which the players engage after player 1 chooses *Concert*. In the second subgame, the terminal histories are (B, B), (B, S), (S, B), and (S, S), the player function assigns the set $\{1, 2\}$ consisting of both players to the empty history (the only nonterminal history), the set of actions of each player at the empty history is $\{B, S\}$, and the players' preferences are represented by the payoffs in the table in Figure 208.1. (This subgame models the same situation as *BoS*.)

A subgame perfect equilibrium is defined as before: a *subgame perfect equilibrium* of an extensive game with perfect information and simultaneous moves is a strategy profile with the property that in no subgame can any player increase her payoff by choosing a different strategy, given the other players' strategies. The formal definition differs from that of a subgame perfect equilibrium of a game without simultaneous moves (Definition 166.1) only in that the meaning of "it is player i's turn to move" is that i is a member of $P(h)$, rather than $P(h) = i$.

To find the set of subgame perfect equilibria of an extensive game with perfect information and simultaneous moves that has a finite horizon, we can, as before, use backward induction. The only wrinkle is that some (perhaps all) of the situations we need to analyze are not single-person decision problems, as they are in the absence of simultaneous moves, but problems in which several players choose actions simultaneously. We cannot simply find an optimal action for the player whose turn it is to move at the start of each subgame, given the players' behavior in the remainder of the game. We need to find a *list* of actions for the players who move at the start of each subgame, with the property that each player's action is optimal given the other players' simultaneous actions and the players' behavior in the remainder of the game. That is, the argument we need to make is the same as the one we make when finding a Nash equilibrium of a strategic game. This argument may use any of the techniques discussed in Chapter 2: it may check each action profile in turn, it may construct and study the players' best response functions, or it may show directly that an action profile we have obtained by a combination of intuition and trial and error is an equilibrium.

EXAMPLE 210.1 (Subgame perfect equilibria of a variant of *BoS*) Consider the game in Figure 208.1. Backward induction proceeds as follows.

- In the subgame that follows the history *Concert*, there are two Nash equilibria (in pure strategies), namely (S, S) and (B, B), as we found in Section 2.7.2.
- If the outcome in the subgame that follows *Concert* is (S, S), then the optimal choice of player 1 at the start of the game is *Book*.
- If the outcome in the subgame that follows *Concert* is (B, B), then the optimal choice of player 1 at the start of the game is *Concert*.

We conclude that the game has two subgame perfect equilibria: $((Book, S), S)$ and $((Concert, B), B)$.

Every finite extensive game with perfect information has a (pure) subgame perfect equilibrium (Proposition 173.1). The same is not true of a finite extensive game with perfect information and simultaneous moves because, as we know, a finite strategic game (which corresponds to an extensive game with perfect information and simultaneous moves of length 1) may not possess a pure strategy Nash equilibrium. (Consider *Matching Pennies* (Example 19.1).) If you have studied Chapter 4, you know that some strategic games that lack a pure strategy Nash equilibrium have a "mixed strategy Nash equilibrium", in which each player randomizes. The same is true of extensive games with perfect information and simultaneous moves. However, in this chapter I restrict attention almost exclusively to pure strategy equilibria; the only occasion on which mixed strategy Nash equilibrium appears is Exercise 212.1.

EXERCISE 210.2 (Extensive game with simultaneous moves) Find the subgame perfect equilibria of the following game. First player 1 chooses either A or B. After either choice, she and player 2 simultaneously choose actions. If player 1 initially chooses A, then she and player 2 subsequently each choose either C or D; if player 1 chooses B initially, then she and player 2 subsequently each choose either E or F. Among the terminal histories, player 1 prefers $(A, (C, C))$ to $(B, (E, E))$ to $(A, (D, D))$ to $(B, (F, F))$, and prefers all these to $(A, (C, D))$, $(A, (D, C))$, $(B, (E, F))$, and $(B, (F, E))$, between which she is indifferent. Player 2 prefers $(A, (D, D))$ to $(B, (F, F))$ to $(A, (C, C))$ to $(B, (E, E))$, and prefers all these to $(A, (C, D))$, $(A, (D, C))$, $(B, (E, F))$, and $(B, (F, E))$, between which she is indifferent.

EXERCISE 210.3 (Two-period *Prisoner's Dilemma*) Two people simultaneously select actions; each person chooses either Q or F (as in the *Prisoner's Dilemma*). Then they simultaneously select actions again, once again each choosing either Q or F. Each person's preferences are represented by the payoff function that assigns to the terminal history $((W, X), (Y, Z))$ (where each component is either Q or F) a payoff equal to the sum of the person's payoffs to (W, X) and to (Y, Z) in the *Prisoner's Dilemma* given in Figure 15.1. Specify this situation as an extensive game with perfect information and simultaneous moves and find its subgame perfect equilibria.

EXERCISE 211.1 (Timing claims on an investment) An amount of money accumulates; in period t $(= 1, 2, \ldots, T)$ its size is \$$2t$. In each period two people simultaneously decide whether to claim the money. If only one person does so, she gets all the money; if both people do so, they split the money equally; and if neither person does so, both people have the opportunity to do so in the next period. If neither person claims the money in period T, each person obtains \$$T$. Each person cares only about the amount of money she obtains. Formulate this situation as an extensive game with perfect information and simultaneous moves, and find its subgame perfect equilibria). (Start by considering the cases $T = 1$ and $T = 2$.)

EXERCISE 211.2 (A market game) A seller owns one indivisible unit of a good, which she does not value. Several potential buyers, each of whom attaches the same positive value v to the good, simultaneously offer prices they are willing to pay for the good. After receiving the offers, the seller decides which, if any, to accept. If she does not accept any offer, then no transaction takes place, and all payoffs are 0. Otherwise, the buyer whose offer the seller accepts pays the amount p she offered and receives the good; the payoff of the seller is p, the payoff of the buyer who obtained the good is $v - p$, and the payoff of every other buyer is 0. Model this situation as an extensive game with perfect information and simultaneous moves and find its subgame perfect equilibria. (Use a combination of intuition and trial and error to find a strategy profile that appears to be an equilibrium, then argue directly that it is. The incentives in the game are closely related to those in Bertrand's oligopoly game (see Exercise 68.1), with the roles of buyers and sellers reversed.) Show, in particular, that in every subgame perfect equilibrium every buyer's payoff is zero.

MORE EXPERIMENTAL EVIDENCE ON SUBGAME PERFECT EQUILIBRIUM

Experiments conducted in 1989 and 1990 among college students (mainly taking economics classes) show that the subgame perfect equilibria of the game in Exercise 211.2 correspond closely to experimental outcomes (Roth, Prasnikar, Okuno-Fujiwara, and Zamir 1991), in contrast to the subgame perfect equilibrium of the ultimatum game (see the box on page 183).

In experiments conducted at four locations (Jerusalem, Ljubljana, Pittsburgh, and Tokyo), nine "buyers" simultaneously bid for the rough equivalent (in terms of local purchasing power) of U.S.\$10, held by a "seller". Each experiment involved a group of 20 participants, which was divided into two markets, each with one seller and nine buyers. Each participant was involved in ten rounds of the market; in each round the sellers and buyers were assigned anew, and in any given round no participant knew who, among the other participants, were sellers and buyers, and who was involved in her market. In every session of the experiment the maximum proposed price was accepted by the seller, and by the seventh round of every experiment the highest bid was at least (the equivalent of) U.S.$9.95.

Experiments involving the ultimatum game, run in the same locations using a similar design, yielded results similar to those of previous experiments (see the box on page 183): proposers kept considerably less than 100% of the pie, and nontrivial offers were rejected.

The box on page 183 discusses two explanations for the experimental results in the ultimatum game. Both explanations are consistent with the results in the market game. One explanation is that people are concerned not only with their own monetary payoffs, but also with other people's payoffs. At least some specifications of such preferences do not affect the subgame perfect equilibria of a market game with many buyers, which still all yield every buyer the payoff of zero. (When there are many buyers, even a seller who cares about the other players' payoffs accepts the highest price offered, because accepting a lower price has little impact on the distribution of monetary payoffs, all but two of which remain zero.) Thus such preferences are consistent with both sets of experimental outcomes. Another explanation is that people incorrectly recognize the ultimatum game as one in which the rule of thumb "don't be a sucker" is advantageously invoked and thus reject a poor offer, "punishing" the person who makes such an offer. In the market game, the players treated poorly in the subgame perfect equilibrium are the buyers, who move first, and hence have no opportunity to punish any other player. Thus the rule of thumb is not relevant in this game, so that this explanation is also consistent with both sets of experimental outcomes.

In the next exercise you are asked to investigate subgame perfect equilibria in which some players use mixed strategies (discussed in Chapter 4).

EXERCISE 212.1 (Price competition) Extend the model in Exercise 128.1 by having the sellers simultaneously choose their prices before the buyers simultaneously choose which seller to approach. Assume that each seller's preferences are represented by the expected value of a Bernoulli payoff function in which the payoff to not trading is 0 and the payoff to trading at the price p is p. Formulate this model precisely as an extensive game with perfect information and simultaneous moves. Show that for every $p \geq \frac{1}{2}$ the game has a subgame perfect equilibrium in which each seller announces the price p. (You may use the fact that if seller j's price is at least $\frac{1}{2}$, seller i's payoff in the mixed strategy equilibrium of the subgame in which the buyers choose which seller to approach is decreasing in her price p_i when $p_i > p_j$.)

7.2 Illustration: entry into a monopolized industry

7.2.1 General model

An industry is currently monopolized by a single firm (the "incumbent"). A second firm (the "challenger") is considering entry, which entails a positive cost f in addition to its production cost. If the challenger stays out, then its profit is zero, whereas if it enters, the firms simultaneously choose outputs (as in Cournot's model of duopoly (Section 3.1)). The cost to firm i of producing q_i units of output is $C_i(q_i)$. If the firms' total output is Q, then the market price is $P_d(Q)$. (As in Section 6.2, I add a subscript to P to avoid a clash with the player function of the game.)

We can model this situation as the following extensive game with perfect information and simultaneous moves, illustrated in Figure 213.1.

Players The two firms: the incumbent (firm 1) and the challenger (firm 2).

Terminal histories $(In, (q_1, q_2))$ for any pair (q_1, q_2) of outputs (nonnegative numbers), and (Out, q_1) for any output q_1.

Player function $P(\varnothing) = \{2\}$, $P(In) = \{1, 2\}$, and $P(Out) = \{1\}$.

Actions $A_2(\varnothing) = \{In, Out\}$; $A_1(In)$, $A_1(Out)$, and $A_2(In)$ are all equal to the set of possible outputs (nonnegative numbers).

Preferences Each firm's preferences are represented by its profit, which for a terminal history $(In, (q_1, q_2))$ is $q_1 P_d(q_1 + q_2) - C_1(q_1)$ for the incumbent and $q_2 P_d(q_1 + q_2) - C_2(q_2) - f$ for the challenger, and for a terminal history (Out, q_1) is $q_1 P_d(q_1) - C_1(q_1)$ for the incumbent and 0 for the challenger.

7.2.2 Example

Suppose that $C_i(q_i) = cq_i$ for all q_i ("unit cost" is constant, equal to c), and the inverse demand function is linear where it is positive, given by $P_d(Q) = \alpha - Q$ for $Q \leq \alpha$, as in Section 3.1.3. To find the subgame perfect equilibria, first consider the subgame that follows the history In. The strategic form of this subgame is the same as the example of Cournot's duopoly game studied in Section 3.1.3, except that the payoff of the challenger is reduced by f (the fixed cost of entry) regardless of the challenger's output. Thus the subgame has a unique Nash equilibrium, in

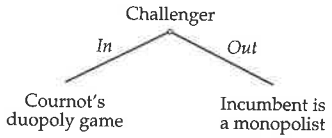


Figure 213.1 An entry game.

which the output of each firm is $\frac{1}{3}(\alpha - c)$; the incumbent's profit is $\frac{1}{9}(\alpha - c)^2$, and the challenger's profit is $\frac{1}{9}(\alpha - c)^2 - f$.

Now consider the subgame that follows the history *Out*. In this subgame the incumbent chooses an output. The incumbent's profit when it chooses the output q_1 is $q_1(\alpha - q_1) - cq_1 = q_1(\alpha - c - q_1)$. This function is a quadratic that increases and then decreases as q_1 increases, and is zero when $q_1 = 0$ and when $q_1 = \alpha - c$. Thus the function is maximized when $q_1 = \frac{1}{2}(\alpha - c)$. We conclude that in any subgame perfect equilibrium the incumbent chooses $q_1 = \frac{1}{2}(\alpha - c)$ in the subgame following the history *Out*.

Finally, consider the challenger's action at the start of the game. If the challenger stays out, then its profit is 0, whereas if it enters, then given the actions chosen in the resulting subgame, its profit is $\frac{1}{9}(\alpha - c)^2 - f$. Thus in any subgame perfect equilibrium the challenger enters if $\frac{1}{9}(\alpha - c)^2 > f$ and stays out if $\frac{1}{9}(\alpha - c)^2 < f$. If $\frac{1}{9}(\alpha - c)^2 = f$, then the game has two subgame perfect equilibria; the challenger enters in one and does not in the other.

In summary, the set of subgame perfect equilibria depends on the value of f. In all equilibria the incumbent's strategy is to produce $\frac{1}{3}(\alpha - c)$ if the challenger enters and $\frac{1}{2}(\alpha - c)$ if it does not, and the challenger's strategy involves its producing $\frac{1}{3}(\alpha - c)$ if it enters.

- If $f < \frac{1}{9}(\alpha - c)^2$ there is a unique subgame perfect equilibrium, in which the challenger enters. The outcome is that the challenger enters and each firm produces the output $\frac{1}{3}(\alpha - c)$.

- If $f > \frac{1}{9}(\alpha - c)^2$ there is a unique subgame perfect equilibrium, in which the challenger stays out. The outcome is that the challenger stays out and the incumbent produces $\frac{1}{2}(\alpha - c)$.

- If $f = \frac{1}{9}(\alpha - c)^2$ the game has two subgame perfect equilibria: the one for the case $f < \frac{1}{9}(\alpha - c)^2$ and the one for the case $f > \frac{1}{9}(\alpha - c)^2$.

Why, if f is small, does the game have no subgame perfect equilibrium in which the incumbent floods the market if the challenger enters, so that the challenger optimally stays out and the incumbent obtains a profit higher than it would have if the challenger had entered? Because the action this strategy prescribes after the history in which the challenger enters is not the incumbent's action in a Nash equilibrium of the subgame: the subgame has a unique Nash equilibrium, in which each firm produces $\frac{1}{3}(\alpha - c)$. Put differently, the incumbent's "threat" to flood the market if the challenger enters is not credible.

? EXERCISE 214.1 (Bertrand's duopoly game with entry) Find the subgame perfect equilibria of the variant of the game studied in this section in which the post-entry competition is a game in which each firm chooses a price, as in the example of Bertrand's duopoly game studied in Section 3.2.2, rather than an output.

7.3 Illustration: electoral competition with strategic voters

The voters in Hotelling's model of electoral competition (Section 3.3) are not players in the game: each citizen is assumed simply to vote for the candidate whose position she most prefers. How do the conclusions of the model change if we assume that each citizen *chooses* the candidate for whom to vote?

Consider the extensive game in which the candidates first simultaneously choose actions, then the citizens simultaneously choose how to vote. As in the variant of Hotelling's game considered on page 74, assume that each candidate may either choose a position (as in Hotelling's original model) or choose to stay out of the race, an option she is assumed to rank between losing and tying for first place with all the other candidates.

Players The candidates and the citizens.

Terminal histories All sequences (x, v) where x is a list of the candidates' actions, each component of which is either a position (a number) or *Out*, and v is a list of voting decisions for the citizens (i.e. a list of candidates, one for each citizen).

Player function $P(\varnothing)$ is the set of all the candidates, and $P(x)$, for any list x of positions for the candidates, is the set of all citizens.

Actions The set of actions available to each candidate at the start of the game consists of *Out* and the set of possible positions. The set of actions available to each citizen after a history x is the set of candidates.

Preferences Each candidate's preferences are represented by a payoff function that assigns n to every terminal history in which she wins outright, k to every terminal history in which she ties for first place with $n - k$ other candidates (for $1 \leq k \leq n - 1$), 0 to every terminal history in which she stays out of the race, and -1 to every terminal history in which she loses, where n is the number of candidates. Each citizen's preferences are represented by a payoff function that assigns to each terminal history the average distance from the citizen's favorite position of the set of winning candidates in that history.

First consider the game in which there are two candidates (and an arbitrary number of citizens). Every subgame following choices of positions by the candidates has many Nash equilibria (as you know if you solved Exercise 48.1). For example, any action profile in which *all* citizens vote for the same candidate is a Nash equilibrium. (A citizen's switching her vote to another candidate has no effect on the outcome.)

This plethora of Nash equilibria allows us to construct, for *every* pair of positions, a subgame perfect equilibrium in which the candidates choose those positions! Consider the strategy profile in which the candidates choose the positions x_1 and x_2, and

- all citizens vote for candidate 1 after a history (x_1', x_2') in which $x_1' = x_1$
- all citizens vote for candidate 2 after a history (x_1', x_2') in which $x_1' \neq x_1$.

The outcome is that the candidates choose the positions x_1 and x_2 and candidate 1 wins. The strategy profile is a subgame perfect equilibrium because for every history (x_1, x_2) the profile of the citizens' actions is a Nash equilibrium, and neither candidate can induce an outcome she prefers by deviating: a deviation by candidate 1 to a position different from x_1 leads her to lose, and a deviation by candidate 2 has no effect on the outcome.

However, most of the Nash equilibria of the voting subgames are fragile (as you know if you solved Exercise 48.1): a citizen's voting for her less preferred candidate is weakly dominated (Definition 46.1) by her voting for her favorite candidate. (A citizen who switches from voting for her less preferred candidate to voting for her favorite candidate either does not affect the outcome (if her favorite candidate was three or more votes behind) or causes her favorite candidate either to tie for first place rather than lose, or to win rather than tie.) Thus in the only Nash equilibrium of a voting subgame in which no citizen uses a weakly dominated action, each citizen votes for the candidate whose position is closest to her favorite position.

Hotelling's model (Section 3.3) *assumes* that each citizen votes for the candidate whose position is closest to her favorite position; in its unique Nash equilibrium, each candidate's position is the median of the citizens' favorite positions. Combining this result with the result of the previous paragraph, we conclude that the game we are studying has only one subgame perfect equilibrium in which no player's strategy is weakly dominated: each candidate chooses the median of the citizens' favorite positions, and for every pair of the candidates' positions, each citizen votes for her favorite candidate.

In the game with three or more candidates, not only do many of the voting subgames have many Nash equilibria, with a variety of outcomes, but restricting to voting strategies that are not weakly dominated does not dramatically affect the set of equilibria: a citizen's only weakly dominated strategy is a vote for her least preferred candidate (see Exercise 49.1).

However, the set of equilibrium outcomes is dramatically restricted by the assumption that each candidate prefers to stay out of the race than to enter and lose, as the next two exercises show. The result in the first exercise is that the game has a subgame perfect equilibrium in which no citizen's strategy is weakly dominated and every candidate enters and chooses as her position the median of the citizens' favorite positions. The result in the second exercise is that under an assumption that makes the citizens averse to ties and an assumption that there exist citizens with extreme preferences, in *every* subgame perfect equilibrium all candidates who enter do so at the median of the citizens' favorite positions. The additional assumptions about the citizens' preferences are much stronger than necessary; they are designed to make the argument relatively easy.

? EXERCISE 216.1 (Electoral competition with strategic voters) Assume that there are $n \geq 3$ candidates and q citizens, where $q \geq 2n$ is odd (so that the median of the voters' favorite positions is well defined) and divisible by n. Show that the game has a subgame perfect equilibrium in which no citizen's strategy is weakly dominated and every candidate enters the race and chooses the median of the citizens' favorite positions. (You may use the fact that every voting subgame has a (pure) Nash equilibrium in which no citizen's action is weakly dominated.)

?? EXERCISE 217.1 (Electoral competition with strategic voters) Consider the variant of the game in this section in which (*i*) the set of possible positions is the set of numbers x with $0 \leq x \leq 1$, (*ii*) the favorite position of at least one citizen is 0 and the favorite position of at least one citizen is 1, and (*iii*) each citizen's preferences are represented by a payoff function that assigns to each terminal history the distance from the citizen's favorite position to the position of the candidate in the set of winners whose position is *furthest* from her favorite position. Under the other assumptions of the previous exercise, show that in every subgame perfect equilibrium in which no citizen's action is weakly dominated, the position chosen by every candidate who enters is the median of the citizens' favorite positions. To do so, first show that in any equilibrium each candidate that enters is in the set of winners. Then show that in any Nash equilibrium of any voting subgame in which there are more than two candidates and not all candidates' positions are the same, some candidate loses. (Argue that if all candidates tie for first place, some citizen can increase her payoff by changing her vote.) Finally, show that in any subgame perfect equilibrium in which either only two candidates enter, or all candidates who enter choose the same position, every entering candidates chooses the median of the citizens' favorite positions.

7.4 Illustration: committee decision-making

How does the procedure used by a committee affect the decision it makes? One approach to this question models a decision-making procedure as an extensive game with perfect information and simultaneous moves in which there is a sequence of ballots, in each of which the committee members vote simultaneously; the result of each ballot determines the choices on the next ballot, or, eventually, the decision to be made.

Fix a set of committee members and a set of *alternatives* over which each member has strict preferences (no member is indifferent between any two alternatives). Assume that the number of committee members is odd, to avoid ties in votes. If there are two alternatives, the simplest committee procedure is that in which the members vote simultaneously for one of the alternatives. (We may interpret the game in Section 2.9.3 as a model of this procedure.) In the procedure illustrated in Figure 218.1, there are three alternatives, x, y, and z. The committee first votes whether to choose x (option "a") or to eliminate it from consideration (option "b"). If it votes to eliminate x, it subsequently votes between y and z.

In these procedures, each vote is between two options. Such procedures are called *binary agendas*. We may define a binary agenda with the aid of an auxiliary one-player extensive game with perfect information in which the set $A(h)$ of ac-

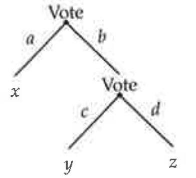


Figure 218.1 A voting procedure, or "binary agenda".

tions following any nonterminal history h has two members, and the number of terminal histories is at least the number of alternatives. We associate with every terminal history h of this auxiliary game an alternative $\alpha(h)$ in such a way that each alternative is associated with at least one terminal history.

In the binary agenda associated with the auxiliary game G, all players vote simultaneously whenever the player in G takes an action. The options on the ballot following the nonterminal history in which a majority of committee members choose option a^1 at the start of the game, then option a^2, and so on, are the members of the set $A(a^1,\ldots,a^k)$ of actions of the player in G after the history $(a^1,\ldots,a^k)$. The alternative selected after the terminal history in which the majority choices are a^1, $\ldots$, a^k is the alternative $\alpha(a^1,\ldots,a^k)$ associated with $(a^1,\ldots,a^k)$ in G. For example, in the auxiliary one-person game that defines the structure of the agenda in Figure 218.1, the single player first chooses a or b; if she chooses a the game ends, whereas if she chooses b, she then chooses between c and d. The alternative x is associated with the terminal history a, y is associated with (b,c), and z is associated with (b,d).

Precisely, the **binary agenda** associated with the auxiliary game G is the extensive game with perfect information and simultaneous moves defined as follows.

Players The set of committee members.

Terminal histories A sequence $(v^1,\ldots,v^k)$ of action profiles (in which each v^j is a list of the players' votes) is a terminal history if and only if there is a terminal history $(a^1,\ldots,a^k)$ of G such that for every $j = 0,\ldots,k-1$, every element of v^{j+1} is a member of $A(a^1,\ldots,a^j)$ ($A(\varnothing)$ if $j = 0$) and a majority of the players' actions in v^{j+1} are equal to a^{j+1}.

Player function For every nonterminal history h, $P(h)$ is the set of all players.

Actions For every player i and every nonterminal history $(v^1,\ldots,v^j)$, player i's set of actions is $A(a^1,\ldots,a^j)$, where $(a^1,\ldots,a^j)$ is the history of G in which, for all ℓ, a^ℓ is the action chosen by the majority of players in v^ℓ.

Preferences The rank each player assigns to the terminal history $(v^1,\ldots,v^k)$ is equal to the rank she assigns to the alternative $\alpha(a^1,\ldots,a^k)$ associated with the terminal history $(a^1,\ldots,a^k)$ of G in which, for all j, a^j is the action chosen by a majority of players in v^j.

Every binary agenda, like every voting subgame of the model in the previous section, has many subgame perfect equilibria. In fact, in any binary agenda, *every* alternative is the outcome of some subgame perfect equilibrium, because if, in every vote, every player votes for the same option, no player can affect the outcome by changing her strategy. However, if we restrict attention to weakly undominated strategies, we greatly reduce the set of equilibria. As we saw before (Section 2.9.3), in a ballot with two options, a player's action of voting for the option she prefers weakly dominates the action of voting for the other option. Thus in a subgame perfect equilibrium of a binary agenda in which every player's vote on every ballot is weakly undominated, on each ballot every player votes for the option that leads, ultimately (given the outcomes of the later ballots), to the alternative she prefers. The alternative associated with the terminal history generated by such a subgame perfect equilibrium is said to be the outcome of *sophisticated voting*.

Which alternatives are the outcomes of sophisticated voting in binary agendas? Say that alternative x *beats* alternative y if a majority of committee members prefer x to y. An alternative that beats every other alternative is called a *Condorcet winner*. For any preferences, there is either one Condorcet winner or no Condorcet winner (see Exercise 75.3).

First suppose that the players' preferences are such that some alternative, say x^*, is a Condorcet winner. I claim that x^* is the outcome of sophisticated voting in *every* binary agenda. The argument, using backward induction, is simple. First consider a subgame of length 1 in which one option leads to the alternative x^*. In this subgame a majority of the players vote for the option that leads to x^*, because a majority prefers x^* to every other alternative, and each player's only weakly undominated strategy is to vote for the option that leads to the alternative she prefers. Thus in at least one subgame of length 2, at least one option leads ultimately to the decision x^* (given the players' votes in the subgames of length 1). In this subgame, by the same argument as before, the winning option leads to x^*. Continuing backward, we conclude that at least one option on the first ballot leads ultimately to x^* and that consequently the winning option on this ballot leads to x^*.

Thus if the players' preferences are such that a Condorcet winner exists, the agenda does not matter: the outcome of sophisticated voting is always the Condorcet winner. If the players' preferences are such that no alternative is a Condorcet winner, the outcome of sophisticated voting depends on the agenda. Consider, for example, a committee with three members facing three alternatives. Suppose that one member prefers x to y to z, another prefers y to z to x, and the third prefers z to x to y. For these preferences, no alternative is a Condorcet winner. The outcome of sophisticated voting in the binary agenda in Figure 218.1 is the alternative x. (Use backward induction: y beats z, and x beats y.) If the positions of x and y are interchanged, then the outcome is y, and if the positions of x and z are interchanged, then the outcome is z. Thus in this case, for *every* alternative there is a binary agenda for which that alternative is the outcome of sophisticated voting.

Which alternatives are the outcomes of sophisticated voting in binary agendas when no alternative is a Condorcet winner? Consider a committee with arbitrary preferences (not necessarily the ones considered in the previous paragraph) that uses the agenda in Figure 218.1. For x to be the outcome of sophisticated voting it must beat the winner of y and z. It may not beat both y and z directly, but it must beat them both at least "indirectly": either x beats y beats z, or x beats z beats y. Similarly, if y or z is the outcome of sophisticated voting, then it must beat both of the other alternatives at least indirectly.

Precisely, say that alternative x *indirectly beats* alternative y if for some $k \geq 1$ there are alternatives $u_1, \ldots, u_k$ such that x beats u_1, u_j beats u_{j+1} for $j = 1, \ldots, k-1$, and u_k beats y. The set of alternatives x such that x beats every other alternative either directly or indirectly is called the *top cycle set*. (Note that if alternative x beats any alternative indirectly, it beats at least one alternative directly.) If there is a Condorcet winner, then the top cycle set consists of this single alternative. If there is no Condorcet winner, then the top cycle set contains more than one alternative.

? EXERCISE 220.1 (Top cycle set) A committee has three members.

a. Suppose that there are three alternatives, x, y, and z, and that one member prefers x to y to z, another prefers y to z to x, and the third prefers z to x to y. Find the top cycle set.

b. Suppose that there are four alternatives, w, x, y, and z, and that one member prefers w to z to x to y, one member prefers y to w to z to x, and one member prefers x to y to w to z. Find the top cycle set. Show, in particular, that z is in the top cycle set even though *all* committee members prefer w.

Rephrasing my conclusion for the agenda in Figure 218.1, if an alternative is the outcome of sophisticated voting, then it is in the top cycle set. The argument for this conclusion extends to any binary agenda. In every subgame, the outcome of sophisticated voting must beat the alternative that will be selected if it is rejected. Thus by backward induction, the outcome of sophisticated voting in the whole game must beat every other alternative either directly or indirectly: the outcome of sophisticated voting in any binary agenda is in the top cycle set.

Now consider a converse question: for any given alternative x in the top cycle set, is there a binary agenda for which x is the outcome of sophisticated voting? The answer is affirmative. The idea behind the construction of an appropriate agenda is illustrated by a simple example. Suppose that there are three alternatives, x, y, and z, and x beats y beats z. Then the agenda in Figure 218.1 is one for which x is the outcome of sophisticated voting. Now suppose there are two additional alternatives, u and w, and x beats u beats w. Then we can construct a larger agenda in which x is the outcome of sophisticated voting by replacing the alternative x in Figure 218.1 with a subgame in which a vote is taken for or against x, and, if x is rejected, a vote is subsequently taken between u and w. If there are other chains through which x beats other alternatives, we can similarly add further subgames.

? EXERCISE 221.1 (Designing agendas) A committee has three members; there are five alternatives. One member prefers x to y to v to w to z, another prefers z to x to v to w to y, and the third prefers y to z to w to v to x. Find the top cycle set, and for each alternative a in the set design a binary agenda for which a is the outcome of sophisticated voting. Convince yourself that for no binary agenda is the outcome of sophisticated voting outside the top cycle set.

? EXERCISE 221.2 (An agenda that yields an undesirable outcome) Design a binary agenda for the committee in Exercise 220.1*b* for which the outcome of sophisticated voting is z (which is worse for all committee members than w).

In summary, (*i*) for any binary agenda, the alternative generated by the subgame perfect equilibrium in which no citizen's action in any ballot is weakly dominated is in the top cycle set, and (*ii*) for every alternative in the top cycle set, there is a binary agenda for which that alternative is generated by the subgame perfect equilibrium in which no citizen's action in any ballot is weakly dominated. In particular, the extent to which the procedure used by a committee affects its decision depends on the nature of the members' preferences. At one extreme, for preferences such that some alternative is a Condorcet winner, the agenda is irrelevant. At another extreme, for preferences for which every alternative is in the top cycle set, the agenda is instrumental in determining the decision. Further, for some preferences there are agendas for which the subgame perfect equilibrium yields an alternative that is unambiguously undesirable in the sense that there is another alternative that *all* committee members prefer.

7.5 Illustration: exit from a declining industry

An industry currently consists of two firms, one with a large capacity, and one with a small capacity. Demand for the firms' output is declining steadily over time. When will the firms leave the industry? Which firm will leave first? Do the firms' financial resources affect the outcome? The analysis of a model that answers these questions illustrates a use of backward induction more sophisticated than that in the previous sections of this chapter.

7.5.1 A model

Take time to be a discrete variable, starting in period 1. Denote by $P_t(Q)$ the market price in period t when the firms' total output is Q, and assume that this price is declining over time: for every value of Q, we have $P_{t+1}(Q) < P_t(Q)$ for all $t \geq 1$. (See Figure 223.1.) We are interested in the firms' decisions to exit, rather than their decisions of how much to produce in the event they stay in the market, so we assume that firm i's only decision is whether to produce some fixed output, denoted k_i, or to produce no output. (You may think of k_i as firm i's capacity.) Once a firm has stopped production, it cannot start up again. Assume that $k_2 < k_1$ (firm 2 is smaller than firm 1) and that each firm's cost of producing q units of output is cq.

The following extensive game with simultaneous moves models this situation.

Players The two firms.

Terminal histories All sequences $(X^1, \ldots, X^t)$ for some $t \geq 1$, where $X^s = (Stay, Stay)$ for $1 \leq s \leq t-1$ and $X^t = (Exit, Exit)$ (both firms exit in period t), or $X^s = (Stay, Stay)$ for all s with $1 \leq s \leq r-1$ for some r, $X^r = (Stay, Exit)$ or $(Exit, Stay)$, $X^s = Stay$ for all s with $r+1 \leq s \leq t-1$, and $X^t = Exit$ (one firm exits in period r and the other exits in period t), and all infinite sequences $(X^1, X^2, \ldots)$, where $X^r = (Stay, Stay)$ for all r (neither firm ever exits).

Player function $P(h) = \{1,2\}$ after any history h in which neither firm has exited; $P(h) = 1$ after any history h in which only firm 2 has exited; and $P(h) = 2$ after any history h in which only firm 1 has exited.

Actions Whenever a firm moves, its set of actions is $\{Stay, Exit\}$.

Preferences Each firm's preferences are represented by a payoff function that associates with each terminal history the firm's total profit, where the profit of firm i $(= 1, 2)$ in period t is $(P_t(k_i) - c)k_i$ if the other firm has exited and $(P_t(k_1 + k_2) - c)k_i$ if the other firm has not exited.

7.5.2 *Subgame perfect equilibrium*

In a period in which $P_t(k_i) < c$, firm i makes a loss even if it is the only firm remaining (the market price for its output is less than its unit cost). Denote by t_i the last period in which firm i is profitable if it is the only firm in the market. That is, t_i is the largest value of t for which $P_t(k_i) \geq c$. (Refer to Figure 223.1.) Because $k_1 > k_2$, we have $t_1 \leq t_2$: the time at which the large firm becomes unprofitable as a loner is no later than the time at which the small firm becomes unprofitable as a loner.

The game has an infinite horizon, but after period t_i firm i's profit is negative even if it is the only firm remaining in the market. Thus if firm i is in the market in any period after t_i, it chooses *Exit* in that period in every subgame perfect equilibrium. In particular, both firms choose *Exit* in every period after t_2. We can use backward induction from period t_2 to find the firms' subgame perfect equilibrium actions in earlier periods.

If firm 1 (the larger firm) is in the market in any period from t_1 on, it should exit, regardless of whether firm 2 is still operating. As a consequence, if firm 2 is still operating in any period from $t_1 + 1$ to t_2 it should stay: firm 1 will exit in any such period, and in its absence firm 2's profit is positive.

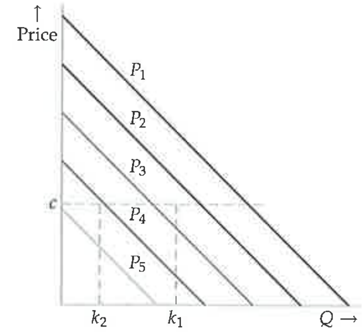


Figure 223.1 The inverse demand curves in a declining industry. In this example, t_1 (the last period in which firm 1 is profitable if it is the only firm in the market) is 2, and t_2 is 4.

So far we have concluded that in every subgame perfect equilibrium, firm 1's strategy is to exit in every period from $t_1 + 1$ on if it has not already done so, and firm 2's strategy is to exit in every period from $t_2 + 1$ on if it has not already done so.

Now consider period t_1, the last period in which firm 1's profit is positive if firm 2 is absent. If firm 2 exits, its profit from then on is zero. If it stays and firm 1 exits, then it earns a profit from period t_1 to period t_2, after which it leaves. If both firms stay, firm 2 sustains a loss in period t_1 but earns a profit in the subsequent periods up to t_2, because in every subgame perfect equilibrium firm 1 exits in period $t_1 + 1$. Thus if firm 2's one-period loss in period t_1 when firm 1 stays in that period is less than the sum of its profits from period $t_1 + 1$ on, then *regardless of whether firm 1 stays or exits in period* t_1, firm 2 stays in every subgame perfect equilibrium. In period $t_1 + 1$, when firm 1 is absent from the industry, the price is relatively high, so that the assumption that firm 2's one-period loss is less than its subsequent multiperiod profit is valid for a significant range of parameters. From now on, I assume that this condition holds.

We conclude that in every subgame perfect equilibrium firm 2 stays in period t_1, so that firm 1 optimally exits. (It definitely exits in the next period, and if it stays in period t_1 it makes a loss, because firm 2 stays.)

Now continue to work backward. If firm 2 stays in period $t_1 - 1$ it earns a profit in periods t_1 through t_2, because in every subgame perfect equilibrium firm 1 exits in period t_1. It may make a loss in period $t_1 - 1$ (if firm 1 stays in that period), but this loss is less than the loss it makes in period t_1 in the company of firm 1, which we have assumed is outweighed by its subsequent profit. Thus regardless of firm 1's action in period $t_1 - 1$, firm 2's best action is to stay in that period. If $t_2 < t_1 - 1$, then firm 1 makes a loss in period $t_1 - 1$ in the company of firm 2, and so should exit.

The same logic applies to all periods back to the first period in which the firms cannot profitably coexist in the industry: in every such period, in every subgame perfect equilibrium firm 1 exits if it has not already done so. Denote by t_0 the last period in which both firms can profitably coexist in the industry: that is, t_0 is the largest value of t for which $P_t(k_1 + k_2) \geq c$.

We conclude that if firm 2's loss in period t_1 when both firms are active is less than the sum of its profits in periods $t_1 + 1$ through t_2 when it alone is active, then the game has a unique subgame perfect equilibrium, in which the large firm exits in period $t_0 + 1$, the first period in which both firms cannot profitably coexist in the industry, and the small firm continues operating until period t_2, after which it alone becomes unprofitable.

EXERCISE 224.1 (Exit from a declining industry) Assume that $c = 10$, $k_1 = 40$, $k_2 = 20$, and $P_t(Q) = 100 - t - Q$ for all values of t and Q for which $100 - t - Q > 0$, otherwise $P_t(Q) = 0$. Find the values of t_1 and t_2 and check whether firm 2's loss in period t_1 when both firms are active is less than the sum of its profits in periods $t_1 + 1$ through t_2 when it alone is active.

7.5.3 *The effect of a constraint on firm 2's debt*

When the firms follow their subgame perfect equilibrium strategies, each firm's profit is nonnegative in every period. However, the equilibrium depends on firm 2's ability to go into debt. Firm 2's strategy calls for it to stay in the market if firm 1, contrary to its strategy, does not exit in the first period in which the market cannot profitably sustain both firms. This feature of firm 2's strategy is essential to the equilibrium. If such a deviation by firm 1 induces firm 2 to exit, then firm 1's strategy of exiting may not be optimal, and the equilibrium may consequently fall apart.

Consider an extreme case, in which firm 2 can never go into debt. We can incorporate this assumption into the model by making firm 2's payoff a large negative number for any terminal history in which its profit in any period is negative. (The size of firm 2's profit depends on the contemporaneous action of firm 1, so we cannot easily incorporate the assumption by modifying the choices available to firm 2.) Consider a history in which firm 1 stays in the market after the last period in which the market can profitably sustain both firms. After such a history firm 2's best action is no longer to stay: if it does so its profit is negative, whereas if it exits its profit is zero. Thus if firm 1 deviates from its equilibrium strategy in the absence of a borrowing constraint for firm 2, and stays in the first period in which it is supposed to exit, then firm 2 optimally exits, and firm 1 reaps positive profits for several periods, as the lone firm in the market. Consequently in this case firm 2 exits first; firm 1 stays in the market until period t_1.

How much debt does firm 2 need to be able to bear for the game to have a subgame perfect equilibrium in which firm 1 exits in period t_0 and firm 2 stays until period t_2? Suppose that firm 2 can sustain losses from period $t_0 + 1$ through period $t_0 + k$, but no longer, when both firms stay in the market. For firm 1 to optimally exit in period $t_0 + 1$, the consequence of its staying in the market must be that firm 2 also stays. Suppose that firm 2's strategy is to stay through period $t_0 + k$, but no longer, if firm 1 does so. Which strategy is best for firm 1 in the subgame starting in period $t_0 + 1$? If it exits, its payoff is zero. If it stays through period $t_0 + k$, its payoff is negative (it makes a loss in every period). If it stays beyond period $t_0 + k$ (when firm 2 exits), it should stay until period t_1, when its payoff is the sum of profits that are negative from period $t_0 + 1$ through period $t_0 + k$ and then positive through period t_1. (See Figure 225.1.) If this payoff is positive it should stay through period t_1; otherwise it should exit immediately.

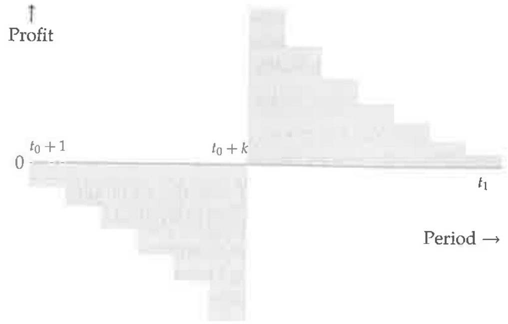


Figure 225.1 Firm 1's profits starting in period $t_0 + 1$ when firm 2 stays in the market until period $t_1 + k$ and firm 1 stays until period t_1.

We conclude that for firm 1 to exit in period $t_0 + 1$, the period until which firm 2 can sustain losses, which I have denoted $t_0 + k$, must be large enough that firm 1's total profit is nonpositive from period $t_0 + 1$ through period t_1 if it shares the market with firm 2 until period $t_0 + k$ and then has the market to itself. This value of k determines the debt that firm 2 must be able to accumulate: the requisite debt equals its total loss when it remains in the market with firm 1 from period $t_0 + 1$ through period $t_0 + k$.

EXERCISE 225.1 (Effect of borrowing constraint of firms' exit decisions in declining industry) Under the assumptions of Exercise 224.1, how much debt does firm 2 need to be able to bear for the subgame perfect equilibrium outcome in the absence of a debt constraint to remain a subgame perfect equilibrium outcome?

7.6 Allowing for exogenous uncertainty

7.6.1 *General model*

The model of an extensive game with perfect information (with or without simultaneous moves) does not allow random events to occur during the course of play.

However, we can easily extend the model to cover such situations. The definition of an **extensive game with perfect information and chance moves** is a variant of the definition of an extensive game with perfect information (155.1) in which

- the player function assigns "chance", rather than a set of players, to some histories
- the probabilities that chance uses after any such history are specified
- the players' preferences are defined over the set of lotteries over terminal histories (rather than simply over the set of terminal histories).

(We may similarly add chance moves to an extensive game with perfect information and simultaneous moves by modifying Definition 206.1.) To keep the analysis simple, assume that the random event after any given history is independent of the random event after any other history. (That is, the realization of any random event is not affected by the realization of any other random event.)

The definition of a player's strategy remains the same as before. The outcome of a strategy profile is now a probability distribution over terminal histories. The definition of subgame perfect equilibrium remains the same as before.

◆ EXAMPLE 226.1 (Extensive game with chance moves) Consider a situation involving two players in which player 1 first chooses A or B. If she chooses A the game ends, with (Bernoulli) payoffs $(1,1)$. If she chooses B, then with probability $\frac{1}{2}$ the game ends, with payoffs $(3,0)$, and with probability $\frac{1}{2}$ player 2 gets to choose between C, which yields payoffs $(0,1)$ and D, which yields payoffs $(1,0)$. An extensive game with perfect information and chance moves that models this situation is shown in Figure 226.1. The label c denotes chance; the number beside each action of chance is the probability with which that action is chosen.

We may use backward induction to find the subgame perfect equilibria of this game. In any equilibrium, player 2 chooses C. Now consider the consequences of player 1's actions. If she chooses A, then she obtains the payoff 1. If she chooses B, then she obtains 3 with probability $\frac{1}{2}$ and 0 with probability $\frac{1}{2}$, yielding an expected payoff of $\frac{3}{2}$. Thus the game has a unique subgame perfect equilibrium, in which player 1 choose B and player 2 chooses C.

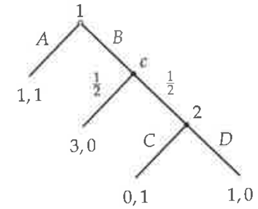


Figure 226.1 An extensive game with perfect information and chance moves. The label c denotes chance; the number beside each action of chance is the probability with which that action is chosen.

? EXERCISE 227.1 (Variant of ultimatum game with equity-conscious players) Consider a variant of the game in Exercise 183.4 in which $\beta_1 = 0$, and the person 2 whom person 1 faces is drawn randomly from a population in which the fraction p have $\beta_2 = 0$ and the remaining fraction $1 - p$ have $\beta_2 = 1$. When making her offer, person 1 knows only that her opponent's characteristic is $\beta_2 = 0$ with probability p and $\beta_2 = 1$ with probability $1 - p$. Model this situation as an extensive game with perfect information and chance moves in which person 1 makes an offer, then chance determines the type of person 2, and finally person 2 accepts or rejects person 1's offer. Find the subgame perfect equilibria of this game. (Use the fact that if $\beta_2 = 0$, then in any subgame perfect equilibrium of the game in Exercise 183.4, person 2 accepts all offers $x > 0$ and may accept or reject the offer 0, and if $\beta_2 = 1$, then she accepts all offers $x > \frac{1}{3}$, may accept or reject the offer $\frac{1}{3}$, and rejects all offers $x < \frac{1}{3}$.) Are there any values of p for which an offer is rejected in equilibrium?

? EXERCISE 227.2 (Firm–union bargaining) A firm knows the amount by which its revenue exceeds its outlays on plant and equipment, but a union with which it is bargaining does not. This "surplus", which is to be divided between the firm and the union, is H with probability p and $L < H$ with probability $1 - p$. The bargaining procedure is that of the ultimatum game: the union makes a demand, which the firm either accepts or rejects. We may model this situation as an extensive game in which, between the union's demand and the firm's response, a move of chance determines whether the surplus is H or L. If the size of the surplus is z and the union's demand is x, then the firm's (Bernoulli) payoff is $z - x$ and the union's (Bernoulli) payoff is x if the firm accepts the union's demand and each player's payoff is 0 (there is a strike) if the firm rejects the union's demand. Find the subgame perfect equilibria of this game for each possible value of p. Find the probability of a strike for each equilibrium.

? EXERCISE 227.3 (Sequential duel) In a sequential duel, two people alternately have the opportunity to shoot each other; each has an infinite supply of bullets. On each of her turns, a person may shoot or refrain from doing so. Each of person i's shots hits (and kills) its intended target with probability p_i (independently of whether any other shots hit their targets). (If you prefer to think about a less violent situation, interpret the players as political candidates who alternately may launch attacks, which may not be successful, against each other.) Each person cares only about her probability of survival (not about the other person's survival). Model this situation as an extensive game with perfect information and chance moves. Show that the strategy pairs in which neither person ever shoots and in which each person always shoots are both subgame perfect equilibria. (Note that the game does not have a finite horizon, so backward induction cannot be used.)

?? EXERCISE 227.4 (Sequential truel) Each of persons A, B, and C has a gun containing a single bullet. Each person, as long as she is alive, may shoot at any surviving person. First A can shoot, then B (if still alive), then C (if still alive). (As in the previous exercise, you may interpret the players as political candidates. In this exercise, each candidate has a budget sufficient to launch a negative campaign to discredit exactly one of its rivals.) Denote by p_i the probability that player i hits her intended target; assume that $0 < p_i < 1$. Assume that each player wishes to maximize her probability of survival; among outcomes in which her survival probability is the same, she wants the danger posed by any other survivors to be as small as possible. (The last assumption is intended to capture the idea that there is some chance that further rounds of shooting may occur, though the possibility of such rounds is not incorporated explicitly into the game.) Model this situation as an extensive game with perfect information and chance moves. (Draw a diagram. Note that the subgames following histories in which A misses her intended target are the same.) Find the subgame perfect equilibria of the game. (Consider only cases in which p_A, p_B, and p_C are all different.) Explain the logic behind A's equilibrium action. Show that "weakness is strength" for C: she is better off if $p_C < p_B$ than if $p_C > p_B$.

Now consider the variant in which each player, on her turn, has the additional option of shooting into the air (in which case she uses a bullet but does not hit anyone). Find the subgame perfect equilibria of this game when $p_A < p_B$. Explain the logic behind A's equilibrium action.

● EXERCISE 228.1 (Cohesion in legislatures) The following pair of games is designed to study the implications of different legislative procedures for the cohesion of a governing coalition. In both games a legislature consists of three members. Initially a governing coalition, consisting of two of the legislators, is given. There are two periods. At the start of each period a member of the governing coalition is randomly chosen (i.e. each legislator is chosen with probability $\frac{1}{2}$) to propose a bill, which is a partition of one unit of payoff between the three legislators. Then the legislators simultaneously cast votes; each legislator votes either for or against the bill. If two or more legislators vote for the bill, it is accepted. Otherwise the course of events differs between the two games. In a game that models the current U.S. legislature, rejection of a bill in period t leads to a given partition d^t of the pie, where $0 < d_i^t < \frac{1}{2}$ for $i = 1, 2, 3$; the governing coalition (the set from which the proposer of a bill is drawn) remains the same in period 2 following a rejection in period 1. In a game that models the current U.K. legislature, rejection of a bill brings down the government; a new governing coalition is determined randomly, and no legislator receives any payoff in that period. Specify each game precisely and find its subgame perfect equilibrium outcomes. Study the degree to which the governing coalition is cohesive (i.e. all its members vote in the same way).

7.6.2 *Using chance moves to model mistakes*

A game with chance moves may be used to model the possibility that players make mistakes. Suppose, for example, that two people simultaneously choose actions.

	A	B
A	1,1	0,0
B	0,0	0,0

Figure 229.1 The players' Bernoulli payoffs to the four pairs of actions in the game studied in Section 7.6.2.

Each person may choose either A or B. Absent the possibility of mistakes, suppose that the situation is modeled by the strategic game in Figure 229.1, in which the numbers in the boxes are Bernoulli payoffs. This game has two Nash equilibria, (A, A) and (B, B).

Now suppose that each person may make a mistake. With probability $1 - p_i > \frac{1}{2}$ the action chosen by person i is the one she intends, and with probability $p_i < \frac{1}{2}$ it is her other action. We can model this situation as the following extensive game with perfect information, simultaneous moves, and chance moves.

Players The two people.

Terminal histories All sequences of the form $((W, X), Y, Z)$, where W, X, Y, and Z are all either A or B; in the history $((W, X), Y, Z)$ player 1 chooses W, player 2 chooses X, and then chance chooses Y for player 1 and Z for player 2.

Player function $P(\varnothing) = \{1, 2\}$ (both players move simultaneously at the start of the game), and $P(W, X) = P((W, X), Y) = \{c\}$ (chance moves twice after the players have acted, first selecting player 1's action and then player 2's action).

Actions The set of actions available to each player at the start of the game, and to chance at each of its moves, is $\{A, B\}$.

Chance probabilities After any history (W, X), chance chooses W with probability $1 - p_1$ and player 1's other action with probability p_1. After any history $((W, X), Y)$, chance chooses X with probability $1 - p_2$ and player 2's other action with probability p_2.

Preferences Each player's preferences are represented by the expected value of a Bernoulli payoff function that assigns 1 to any history $((W, X), A, A)$ (in which chance chooses the action A for each player), and 0 to any other history.

The players in this game move simultaneously, so that the subgame perfect equilibria of the game are its Nash equilibria. To find the Nash equilibria, we construct the strategic form of the game. Suppose that each player chooses the action A. Then the outcome is (A, A) with probability $(1 - p_1)(1 - p_2)$ (the probability that neither player makes a mistake). Thus each player's expected payoff is $(1 - p_1)(1 - p_2)$. Similarly, if player 1 chooses A and player 2 chooses B, then the

	A	B
A	$(1-p_1)(1-p_2), (1-p_1)(1-p_2)$	$(1-p_1)p_2, (1-p_1)p_2$
B	$p_1(1-p_2), p_1(1-p_2)$	p_1p_2, p_1p_2

Figure 230.1 The strategic form of the extensive game with chance moves that models the situation in which with probability p_i each player i in the game in Figure 229.1 chooses an action different from the one she intends.

outcome is (A, A) with probability $(1-p_1)p_2$ (the probability that player 1 does not make a mistake, whereas player 2 does). Making similar computations for the other two cases yields the strategic form in Figure 230.1.

For $p_1 = p_2 = 0$, this game is the same as the original game (Figure 229.1); it has two Nash equilibria, (A, A) and (B, B). If at least one of the probabilities is positive then only (A, A) is a Nash equilibrium: if $p_i > 0$, then $(1-p_j)p_i > p_jp_i$ (given that each probability is less than $\frac{1}{2}$). That is, only the equilibrium (A, A) of the original game is robust to the possibility that the players make small mistakes.

In the original game, each player's action B is weakly dominated (Definition 46.1). Introducing the possibility of mistakes captures the fragility of the equilibrium (B, B): B is optimal for a player only if she is absolutely certain that the other player will choose B also. The slightest chance that the other player will choose A is enough to make A unambiguously the best choice.

We may use the idea that an equilibrium should survive when the players may make small mistakes to discriminate among the Nash equilibria of any strategic game. For two-player games in which each player has finitely many actions, the equilibria that satisfy this requirement are precisely those in which no player's action is weakly dominated. For games with more than two players, no equilibrium in which any player's action is weakly dominated satisfies the requirement, but equilibria in which no player's action is weakly dominated may fail to satisfy the requirement, as the following exercise shows.

? EXERCISE 230.1 (Nash equilibria when players may make mistakes) Consider the three-player game in Figure 230.2. Show that (A, A, A) is a Nash equilibrium in which no player's action is weakly dominated. Now modify the game by assuming that the outcome of any player i's choosing an action X is that X occurs with probability $1 - p_i$ and the player's other action occurs with probability $p_i > 0$. Show that (A, A, A) is not a Nash equilibrium of the modified game when $p_i < \frac{1}{2}$ for $i = 1, 2, 3$.

	A	B
A	1, 1, 1	0, 0, 1
B	1, 1, 1	1, 0, 1

A

	A	B
A	0, 1, 0	1, 0, 0
B	1, 1, 0	0, 0, 0

B

Figure 230.2 A three-player strategic game in which each player has two actions. Player 1 chooses a row, player 2 chooses a column, and player 3 chooses a table.

7.7 Discussion: subgame perfect equilibrium and backward induction

Some of the situations we have studied do not fit well into the idealized setting for the steady state interpretation of a subgame perfect equilibrium discussed in Section 5.4.4, in which each player repeatedly engages in the same game with a variety of randomly selected opponents. In some cases an alternative interpretation fits better: each player deduces her optimal strategy from an analysis of the other players' best actions, given her knowledge of their preferences. Here I discuss a difficulty with this interpretation.

Consider the game in Figure 231.1, in which player 1 moves both before and after player 2. This game has a unique subgame perfect equilibrium, in which player 1's strategy is (B, F) and player 2's strategy is C. Consider player 2's analysis of this game. If she deduces that the only rational action for player 1 at the start of the game is B, then what should she conclude if player 1 chooses A? It seems that she must conclude that something has "gone wrong": perhaps player 1 has made a "mistake", or she misunderstands player 1's preferences, or player 1 is not rational.

If she is convinced that player 1 simply made a mistake, then her analysis of the rest of the game should not be affected. However, if player 1's move induces her to doubt player 1's motivation, she may need to reconsider her analysis of the rest of the game. Suppose, for example, that A and E model similar actions; specifically, suppose that they both correspond to player 1's moving left, whereas B and F both involve her moving right. Then player 1's choice of A at the start of the game may make player 2 wonder whether player 1 confuses left and right, and therefore may choose E after the history (A, C). If so, player 2 should choose D rather than C after player 1 chooses A, giving player 1 an incentive to choose A rather than B at the start of the game.

The next two examples are richer games that more strikingly manifest the difficulty with the alternative interpretation of subgame perfect equilibrium. The first example is an extension of the entry game in Figure 156.1.

◆ EXAMPLE 231.1 (Chain-store game) A chain store operates in K markets. In each market a single challenger must decide whether to compete with it. The chal-

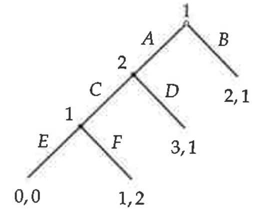


Figure 231.1 An extensive game in which player 1 moves both before and after player 2.

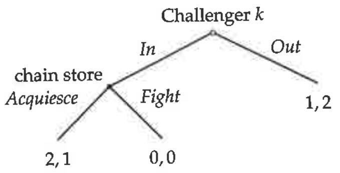


Figure 232.1 The structure of the players' choices in market k in the chain-store game. The first number in each pair is challenger k's profit and the second number is the chain store's profit.

lengers make their decisions sequentially. If any challenger enters, the chain store may acquiesce to its presence (A) or fight it (F). Thus in each period k the outcome is either *Out* (challenger k does not enter), (In, A) (challenger k enters and the chain store acquiesces), or (In, F) (challenger k enters and is fought). When taking an action, any challenger knows all the actions previously chosen. The profits of challenger k and the chain store in market k are shown in Figure 232.1 (cf. Figure 156.1); the chain store's profit in the whole game is the sum of its profits in the K markets.

We can model this situation as the following extensive game with perfect information.

Players The chain store and the K challengers.

Terminal histories The set of all sequences $(e_1, \ldots, e_K)$, where each e_j is either *Out*, (In, A), or (In, F).

Player function The chain store is assigned to every history that ends with *In*, challenger 1 is assigned to the empty history, and challenger k (for $k = 2, \ldots, K$) is assigned to every history $(e_1, \ldots, e_{k-1})$, where each e_j is either *Out*, (In, A), or (In, F).

Preferences Each player's preferences are represented by its profits.

This game has a finite horizon, so we may find its subgame perfect equilibria by using backward induction. Every subgame at the start of which challenger K moves resembles the game in Figure 232.1 for $k = K$; it differs only in that the chain store's profit after each of the three terminal histories is greater by an amount equal to its profit in the previous $K - 1$ markets. Thus in a subgame perfect equilibrium challenger K chooses *In* and the incumbent chooses A in market K.

Now consider the subgame faced by challenger $K - 1$. We know that the outcome in market K is independent of the actions of challenger $K - 1$ and the chain store in market $K - 1$: whatever they do, challenger K enters and the chain store acquiesces to its entry. Thus the chain store should choose its action in market $K - 1$ on the basis of its payoffs in that market alone. We conclude that the chain store's optimal action in market $K - 1$ is A, and challenger $K - 1$'s optimal action is *In*.

We have now concluded that in any subgame perfect equilibrium, the outcome in each of the last two markets is (In, A), regardless of the history. Continuing to work backward to the start of the game we see that the game has a unique subgame perfect equilibrium, in which every challenger enters and the chain store always acquiesces to entry.

● EXERCISE 233.1 (Nash equilibria of chain-store game) Find the set of Nash equilibrium outcomes of the game for an arbitrary value of K. (First think about the case $K = 1$, then generalize your analysis.)

● EXERCISE 233.2 (Subgame perfect equilibrium of chain-store game) Consider the following strategy pair in the game for $K = 100$. For $k = 1, \ldots, 90$, challenger k stays out after any history in which every previous challenger that entered was fought (or no challenger entered), and otherwise enters; challengers 91 through 100 enter. The chain store fights every challenger up to challenger 90 that enters after a history in which it fought every challenger that entered (or no challenger entered), acquiesces to any of these challengers entering after any other history, and acquiesces to challengers 91 through 100 regardless of the history. Find the players' payoffs in this strategy pair. Show that the strategy pair is not a subgame perfect equilibrium: find a player who can increase her payoff in some subgame. By how much can the deviant increase its payoff?

Suppose that $K = 100$. You are in charge of challenger 21. You observe, contrary to the subgame perfect equilibrium, that every previous challenger entered and that the chain store fought each one. What should you do? According to the subgame perfect equilibrium, the chain store will acquiesce to your entry. But should you really regard the chain store's 19 previous decisions as "mistakes"? You might instead read some logic into the chain store's *deliberately* fighting the first 20 entrants: if, by doing so, it persuades more than 20 of the remaining challengers to stay out, then its profit will be higher than it is in the subgame perfect equilibrium. That is, you may imagine that the chain store's aggressive behavior in the earlier markets is an attempt to establish a reputation for being a fighter, which, if successful, will make it better off. By such reasoning you may conclude that your best strategy is to stay out.

Thus, a deviation from the subgame perfect equilibrium by the chain store in which it engages in a long series of fights may not be dismissed by challengers as a series of mistakes, but rather may cause them to doubt the chain store's future behavior. This doubt may lead a challenger who is followed by enough future challengers to stay out.

◆ EXAMPLE 233.3 (Centipede game) The two-player game in Figure 234.1 is known as a "centipede game" because of its shape. (The game, like the arthropod, may have fewer than 100 legs.) The players move alternately; on each move a player can stop the game (S) or continue (C). On any move, a player is better off stopping the game than continuing if the other player stops immediately afterward, but is worse off stopping than continuing if the other player continues, regardless of the subsequent actions. After k periods, the game ends.

This game has a finite horizon, so we may find its subgame perfect equilibria by using backward induction. The last player to move prefers to stop the game

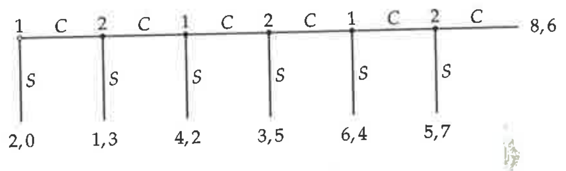


Figure 234.1 A six-period centipede game.

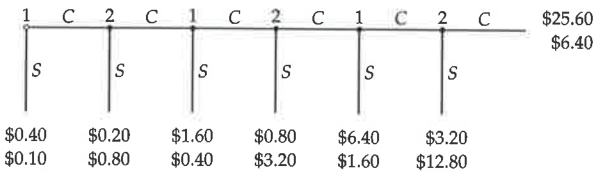
than to continue. Given this player's action, the player who moves before her also prefers to stop the game than to continue. Working backward, we conclude that the game has a unique subgame perfect equilibrium, in which each player's strategy is to stop the game whenever it is her turn to move. The outcome is that player 1 stops the game immediately.

EXERCISE 234.1 (Nash equilibria of the centipede game) Show that the outcome of every Nash equilibrium of this game is the same as the outcome of the unique subgame perfect equilibrium (i.e. player 1 stops the game immediately).

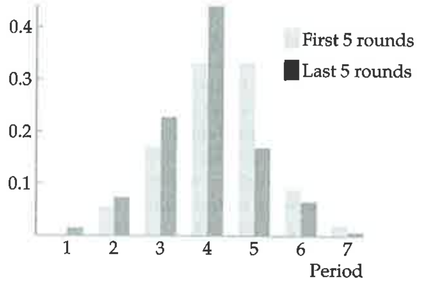
The logic that in the only steady state player 1 stops the game immediately is unassailable. Yet this pattern of behavior is intuitively unappealing, especially if the number *k* of periods is large. The optimality of player 1's choosing to stop the game depends on her believing that if she continues, then player 2 will stop the game in period 2. Further, player 2's decision to stop the game in period 2 depends on her believing that if she continues, then player 1 will stop the game in period 3. Each decision to stop the game is based on similar considerations. Consider a player who has to choose an action in period 21 of a 100-period game, after each player has continued in the first 20 periods. Is she likely to consider the first 20 decisions—half of which were hers—"mistakes"? Or will these decisions induce her to doubt that the other player will stop the game in the next period? These questions have no easy answers; some experimental evidence is discussed in the accompanying box.

EXPERIMENTAL EVIDENCE ON THE CENTIPEDE GAME

In experiments conducted in the United States in 1989, each of 58 student subjects played the game shown below (McKelvey and Palfrey 1992). (The payoff of player 1 is the top amount in each pair.) Each subject played the game 9 or 10 times, facing a different opponent each time; in each play of the game, each subject had previously played the same number of games. Each subject knew in advance how many times she would play the game, and knew that she would not play against the same opponent more than once. If each subject cared only about her own monetary payoff, the game induced by the experiment was a six-period centipede.



The fraction of plays of the game that ended in each period is shown in the graph below. (A game is counted as ending in period 7 if the last player to move chose C. The graph is computed from McKelvey and Palfrey 1992, Table IIIA.) Results are broken down according to the players' experience (first 5 rounds, last 5 rounds). The game ended earlier when the participants were experienced, but even among experienced participants the outcomes are far from the Nash equilibrium outcome, in which the game ends in period 1.



Ten plays of the game may not be enough to achieve convergence to a steady state. But putting aside this limitation of the data, and supposing that convergence was in fact achieved at the end of 10 rounds, how far does the observed behavior differ from a Nash equilibrium (maintaining the assumption that each player cares only about her own monetary payoff)?

The theory of Nash equilibrium has two components: each player optimizes, given her beliefs about the other players, and these beliefs are correct. Some decisions in McKelvey and Palfrey's experiment were patently suboptimal, regardless of the subjects' beliefs: a few subjects in the role of player 2 chose to continue in period 6, obtaining $6.40 with certainty instead of $12.80 with certainty. To assess the departure of the other decisions from optimality we need to assign the subjects beliefs (which were not directly observed). An assumption consistent with the steady state interpretation of Nash equilibrium is that a player's belief is based on her observations of the other players' actions. Even in round 10 of the experiment each player had only nine observations on which to base her belief, and could have used these data in various ways. But suppose that, somehow, at the end of round 4, each player correctly inferred the distribution of her opponents' strategies in the next 5 rounds. What strategy should she subsequently have used? From McKelvey and Palfrey (1992, Table IIIB) we may deduce that the optimal strategy of player 1 stops in period 5 and that of player 2 stops in period 6. That is, each player's best response to the empirical distribution of the other players' strategies differs dramatically from her subgame perfect equilibrium strategy. Other assumptions about the subjects' beliefs rationalize other strategies; the data seem too limited to conclude that the subjects were not optimizing, given the beliefs their experience might reasonably have led them to hold. That is, the experimental data are not strongly inconsistent with the theory of Nash equilibrium as a steady state.

Are the data inconsistent with the theory that rational players, even those with no experience playing the game, will use backward induction to deduce their opponents' rational actions? This theory predicts that the first player immediately stops the game, so certainly the data are inconsistent with it. How inconsistent? One way to approach this question is to consider the implications of each player's thinking that the others are *likely* to be rational, but are not *certainly* so. If, in any period, player 1 thinks that the probability that player 2 will stop the game in the next period is less than $\frac{6}{7}$, continuing yields a higher expected payoff than stopping. Given the limited time the subjects had to analyze the game (and the likelihood that they had never before thought about any related game), even those who understood the implications of backward induction may reasonably have entertained the relatively small doubt about the other players' cognitive abilities required to make stopping the game immediately an unattractive option. Or, alternatively, a player confident of her opponents' logical abilities may have doubted her opponents' assessment of *her own* analytical skills. If player 1 believes that player 2 thinks that the probability that player 1 will continue in period 3 is greater than $\frac{1}{7}$, then she should continue in period 1, because player 2 will continue in period 2. That is, relatively minor departures from the theory yield outcomes close to those observed.

Notes

The idea of regarding games with simultaneous moves as games with perfect information is due to Dubey and Kaneko (1984).

The model in Section 7.3 was first studied by Ledyard (1981, 1984). The approach to voting in committees in Section 7.4 was initiated by Farquharson (1969) (see also Niemi 1983). (The publication of Farquharson's book was delayed; the book was completed in 1958.) The top cycle set was first defined by Ward (1961) (who called it the "majority set"). The characterization of the outcomes of sophisticated voting in binary agendas in terms of the top cycle set is due to Miller (1977) (who calls the top cycle set the "Condorcet set"), McKelvey and Niemi (1978), and Moulin (1986a, pp. 283–284). Miller (1995) surveys the field (see in particular his pp. 85–86). The model in Section 7.5 is taken from Ghemawat and Nalebuff (1985); the idea is closely related to that of Benoît (1984, Section 1) (see Exercise 174.2). My discussion draws on an unpublished exposition of the model by Vijay Krishna. The idea of discriminating among Nash equilibria by considering the possibility that players make mistakes, briefly discussed in Section 7.6.2, is due to Selten (1975). The chain-store game in Example 231.1 is due to Selten (1978). The centipede game in Example 233.3 is due to Rosenthal (1981).

The experimental results discussed in the box on page 211 are due to Roth, Prasnikar, Okuno-Fujiwara, and Zamir (1991). The subgame perfect equilibria of a variant of the market game in which each player's payoff depends on the other players' monetary payoffs are analyzed by Bolton and Ockenfels (2000). The model in Exercise 212.1 is taken from Peters (1984). The results in Exercises 216.1 and 217.1 are due to Feddersen, Sened, and Wright (1990). The game in Exercise 227.4 is a simplification of an example due to Shubik (1954); the main idea appears in Phillips (1937, 159) and Kinnaird (1946, 246), both of which consist mainly of puzzles previously published in newspapers. Exercise 228.1 is based on Diermeier and Feddersen (1998). The experiment discussed in the box on page 234 is reported in McKelvey and Palfrey (1992).