2 Nash Equilibrium: Theory

Prerequisite: Chapter 1.

2.1 Strategic games

A STRATEGIC GAME is a model of interacting decision-makers. In recognition of the interaction, we refer to the decision-makers as *players*. Each player has a set of possible *actions*. The model captures interaction between the players by allowing each player to be affected by the actions of *all* players, not only her own action. Specifically, each player has *preferences* about the action *profile*—the list of all the players' actions. (See Section 17.4, in the mathematical appendix, for a discussion of profiles.)

More precisely, a strategic game is defined as follows. (The qualification "with ordinal preferences" distinguishes this notion of a strategic game from a more general notion studied in Chapter 4.)

▶ DEFINITION 11.1 (*Strategic game with ordinal preferences*) A **strategic game** (with ordinal preferences) consists of

- a set of **players**
- for each player, a set of **actions**
- for each player, **preferences** over the set of action profiles.

A very wide range of situations may be modeled as strategic games. For example, the players may be firms, the actions prices, and the preferences a reflection of the firms' profits. Or the players may be candidates for political office, the actions campaign expenditures, and the preferences a reflection of the candidates' probabilities of winning. Or the players may be animals fighting over some prey, the actions concession times, and the preferences a reflection of whether an animal wins or loses. In this chapter I describe some simple games designed to capture fundamental conflicts present in a variety of situations. The next chapter is devoted to more detailed applications to specific phenomena.

As in the model of rational choice by a single decision-maker (Section 1.2), it is frequently convenient to specify the players' preferences by giving *payoff functions* that represent them. Bear in mind that these payoffs have only *ordinal* significance. If a player's payoffs to the action profiles a, b, and c are 1, 2, and 10, for example, the only conclusion we can draw is that the player prefers c to b and b to a; the numbers do *not* imply that the player's preference between c and b is stronger than her preference between a and b.

Time is absent from the model. The idea is that each player chooses her action once and for all, and the players choose their actions "simultaneously" in the sense that no player is informed, when she chooses her action, of the action chosen by any other player. (For this reason, a strategic game is sometimes referred to as a "simultaneous move game".) Nevertheless, an action may involve activities that extend over time, and may take into account an unlimited number of contingencies. An action might specify, for example, "if company X's stock falls below \$10, buy 100 shares; otherwise, do not buy any shares". (For this reason, an action is sometimes called a "strategy".) However, the fact that time is absent from the model means that when analyzing a situation as a strategic game, we abstract from the complications that may arise if a player is allowed to change her plan as events unfold: we assume that actions are chosen once and for all.

2.2 Example: the *Prisoner's Dilemma*

One of the most well-known strategic games is the *Prisoner's Dilemma*. Its name comes from a story involving suspects in a crime; its importance comes from the huge variety of situations in which the participants face incentives similar to those faced by the suspects in the story.

◆ EXAMPLE 12.1 (*Prisoner's Dilemma*) Two suspects in a major crime are held in separate cells. There is enough evidence to convict each of them of a minor offense, but not enough evidence to convict either of them of the major crime unless one of them acts as an informer against the other (finks). If they both stay quiet, each will be convicted of the minor offense and spend one year in prison. If one and only one of them finks, she will be freed and used as a witness against the other, who will spend four years in prison. If they both fink, each will spend three years in prison.

This situation may be modeled as a strategic game:

Players The two suspects.

Actions Each player's set of actions is $\{Quiet, Fink\}$.

Preferences Suspect 1's ordering of the action profiles, from best to worst, is $(Fink, Quiet)$ (she finks and suspect 2 remains quiet, so she is freed), $(Quiet, Quiet)$ (she gets one year in prison), $(Fink, Fink)$ (she gets three years in prison), $(Quiet, Fink)$ (she gets four years in prison). Suspect 2's ordering is $(Quiet, Fink)$, $(Quiet, Quiet)$, $(Fink, Fink)$, $(Fink, Quiet)$.

We can represent the game compactly in a table. First choose payoff functions that represent the suspects' preference orderings. For suspect 1 we need a function u_1 for which

$$u_1(Fink, Quiet) > u_1(Quiet, Quiet) > u_1(Fink, Fink) > u_1(Quiet, Fink).$$

A simple specification is $u_1(Fink, Quiet) = 3$, $u_1(Quiet, Quiet) = 2$, $u_1(Fink, Fink) = 1$, and $u_1(Quiet, Fink) = 0$. For suspect 2 we can similarly choose the function u_2 for which $u_2(Quiet, Fink) = 3$, $u_2(Quiet, Quiet) = 2$, $u_2(Fink, Fink) = 1$, and $u_2(Fink, Quiet) = 0$. Using these representations, the game is illustrated in Figure 13.1. In this figure the two rows correspond to the two possible actions of player 1, the two columns correspond to the two possible actions of player 2, and the numbers in each box are the players' payoffs to the action profile to which the box corresponds, with player 1's payoff listed first.

		Suspect 2	
		Quiet	*Fink*
Suspect 1	*Quiet*	2, 2	0, 3
	Fink	3, 0	1, 1

Figure 13.1 The *Prisoner's Dilemma* (Example 12.1).

The *Prisoner's Dilemma* models a situation in which there are gains from cooperation (each player prefers that both players choose *Quiet* than they both choose *Fink*) but each player has an incentive to "free ride" (choose *Fink*) whatever the other player does. The game is important not because we are interested in understanding the incentives for prisoners to confess, but because many other situations have similar structures. Whenever each of two players has two actions, say C (corresponding to *Quiet*) and D (corresponding to *Fink*), player 1 prefers (D, C) to (C, C) to (D, D) to (C, D), and player 2 prefers (C, D) to (C, C) to (D, D) to (D, C), the *Prisoner's Dilemma* models the situation that the players face. Some examples follow.

2.2.1 Working on a joint project

You are working with a friend on a joint project. Each of you can either work hard or goof off. If your friend works hard then you prefer to goof off (the outcome of the project would be better if you worked hard too, but the increment in its value to you is not worth the extra effort). You prefer the outcome of your both working hard to the outcome of your both goofing off (in which case nothing gets accomplished), and the worst outcome for you is that you work hard and your friend goofs off (you hate to be "exploited"). If your friend has the same preferences then the game that models the situation you face is given in Figure 14.1, which, as you can see, differs from the *Prisoner's Dilemma* only in the names of the actions.

	Work hard	*Goof off*
Work hard	2, 2	0, 3
Goof off	3, 0	1, 1

Figure 14.1 Working on a joint project.

I am *not* claiming that a situation in which two people pursue a joint project *necessarily* has the structure of the *Prisoner's Dilemma*, only that the players' preferences in such a situation *may* be the same as in the *Prisoner's Dilemma*! If, for example, each person prefers to work hard than to goof off when the other person works hard, then the *Prisoner's Dilemma* does *not* model the situation: the players' preferences are different from those given in Figure 14.1.

? EXERCISE 14.1 (Working on a joint project) Formulate a strategic game that models a situation in which two people work on a joint project in the case that their preferences are the same as those in the game in Figure 14.1 except that each person prefers to work hard than to goof off when the other person works hard. Present your game in a table like the one in Figure 14.1.

2.2.2 Duopoly

In a simple model of a duopoly, two firms produce the same good, for which each firm charges either a low price or a high price. Each firm wants to achieve the highest possible profit. If both firms choose *High* then each earns a profit of $1000. If one firm chooses *High* and the other chooses *Low* then the firm choosing *High* obtains no customers and makes a loss of $200, whereas the firm choosing *Low* earns a profit of $1200 (its unit profit is low, but its volume is high). If both firms choose *Low* then each earns a profit of $600. Each firm cares only about its profit, so we can represent its preferences by the profit it obtains, yielding the game in Figure 14.2.

	High	*Low*
High	1000, 1000	−200, 1200
Low	1200, −200	600, 600

Figure 14.2 A simple model of a price-setting duopoly.

Bearing in mind that what matters are the players' preferences, not the particular payoff functions that we use to represent them, we see that this game, like the previous one, differs from the *Prisoner's Dilemma* only in the names of the actions. The action *High* plays the role of *Quiet*, and the action *Low* plays the role of *Fink*; firm 1 prefers $(Low, High)$ to $(High, High)$ to (Low, Low) to $(High, Low)$, and firm 2 prefers $(High, Low)$ to $(High, High)$ to (Low, Low) to $(Low, High)$.

As in the previous example, I do not claim that the incentives in a duopoly are *necessarily* those in the *Prisoner's Dilemma*; different assumptions about the relative sizes of the profits in the four cases generate a different game. Further, in this case one of the abstractions incorporated into the model—that each firm has only two prices to choose between—may not be harmless; if the firms may choose among many prices then the structure of the interaction may change. (A richer model is studied in Section 3.2.)

2.2.3 The arms race

Under some assumptions about the countries' preferences, an arms race can be modeled as the *Prisoner's Dilemma*. (The *Prisoner's Dilemma* was first studied in the early 1950s, when the USA and USSR were involved in a nuclear arms race, so you might suspect that US nuclear strategy was influenced by game theory; the evidence suggests that it was not.) Assume that each country can build an arsenal of nuclear bombs, or can refrain from doing so. Assume also that each country's favorite outcome is that it has bombs and the other country does not; the next best outcome is that neither country has any bombs; the next best outcome is that both countries have bombs (what matters is relative strength, and bombs are costly to build); and the worst outcome is that only the other country has bombs. In this case the situation is modeled by the *Prisoner's Dilemma*, in which the action *Don't build bombs* corresponds to *Quiet* in Figure 13.1 and the action *Build bombs* corresponds to *Fink*. However, once again the assumptions about preferences necessary for the *Prisoner's Dilemma* to model the situation may not be satisfied: a country may prefer *not* to build bombs if the other country does not, for example (bomb-building may be very costly), in which case the situation is modeled by a different game.

2.2.4 Common property

Two farmers are deciding how much to allow their sheep to graze on the village common. Each farmer prefers that her sheep graze a lot than a little, regardless of the other farmer's action, but prefers that both farmers' sheep graze a little than both farmers' sheep graze a lot (in which case the common is ruined for future use). Under these assumptions the game is the *Prisoner's Dilemma*. (A richer model is studied in Section 3.1.5.)

2.2.5 Other situations modeled as the Prisoner's Dilemma

A huge number of other situations have been modeled as the *Prisoner's Dilemma*, from mating hermaphroditic fish to tariff wars between countries.

? EXERCISE 16.1 (Hermaphroditic fish) Members of some species of hermaphroditic fish choose, in each mating encounter, whether to play the role of a male or a female. Each fish has a preferred role, which uses up fewer resources and hence allows more future mating. A fish obtains a payoff of H if it mates in its preferred role and L if it mates in the other role, where $H > L$. (Payoffs are measured in terms of number of offspring, which fish are evolved to maximize.) Consider an encounter between two fish whose preferred roles are the same. Each fish has two possible actions: mate in either role, and insist on its preferred role. If both fish offer to mate in either role, the roles are assigned randomly, and each fish's payoff is $\frac{1}{2}(H + L)$ (the average of H and L). If each fish insists on its preferred role, the fish do not mate; each goes off in search of another partner, and obtains the payoff S. The higher the chance of meeting another partner, the larger is S. Formulate this situation as a strategic game and determine the range of values of S, for any given values of H and L, for which the game differs from the *Prisoner's Dilemma* only in the names of the actions.

2.3 Example: *Bach or Stravinsky?*

In the *Prisoner's Dilemma* the main issue is whether or not the players will cooperate (choose *Quiet*). In the following game the players agree that it is better to cooperate than not to cooperate, but disagree about the best outcome.

◆ EXAMPLE 16.2 (*Bach or Stravinsky?*) Two people wish to go out together. Two concerts are available: one of music by Bach, and one of music by Stravinsky. One person prefers Bach and the other prefers Stravinsky. If they go to different concerts, each of them is equally unhappy listening to the music of either composer.

We may model this situation as the two-player strategic game in Figure 16.1, in which the person who prefers Bach chooses a row and the person who prefers Stravinsky chooses a column.

	Bach	*Stravinsky*
Bach	2, 1	0, 0
Stravinsky	0, 0	1, 2

Figure 16.1 *Bach or Stravinsky?* (*BoS*) (Example 16.2).

This game is also referred to as the "Battle of the Sexes" (though the conflict it models surely occurs no more frequently between people of the opposite sex than it does between people of the same sex). I call the game *BoS*, an acronym that fits both names. (I assume that each player is indifferent between listening to Bach and listening to Stravinsky when she is alone only for consistency with the standard specification of the game. As we shall see, the analysis of the game remains the same in the absence of this indifference.)

Like the *Prisoner's Dilemma*, *BoS* models a wide variety of situations. Consider, for example, two officials of a political party deciding the stand to take on an issue. Suppose that they disagree about the best stand, but are both better off if they take the same stand than if they take different stands; both cases in which they take different stands, in which case voters do not know what to think, are equally bad. Then *BoS* captures the situation they face. Or consider two merging firms that currently use different computer technologies. As two divisions of a single firm they will both be better off if they both use the same technology; each firm prefers that the common technology be the one it used in the past. *BoS* models the choices the firms face.

2.4 Example: *Matching Pennies*

Aspects of both conflict and cooperation are present in both the *Prisoner's Dilemma* and *BoS*. The next game is purely conflictual.

◆ EXAMPLE 17.1 (*Matching Pennies*) Two people choose, simultaneously, whether to show the Head or the Tail of a coin. If they show the same side, person 2 pays person 1 a dollar; if they show different sides, person 1 pays person 2 a dollar. Each person cares only about the amount of money she receives, and (naturally!) prefers to receive more than less. A strategic game that models this situation is shown in Figure 17.1. (In this representation of the players' preferences, the payoffs are equal to the amounts of money involved. We could equally well work with another representation—for example, 2 could replace each 1, and 1 could replace each -1.)

	Head	*Tail*
Head	$1, -1$	$-1, 1$
Tail	$-1, 1$	$1, -1$

Figure 17.1 *Matching Pennies* (Example 17.1).

In this game the players' interests are diametrically opposed (such a game is called "strictly competitive"): player 1 wants to take the same action as the other player, whereas player 2 wants to take the opposite action.

This game may, for example, model the choices of appearances for new products by an established producer and a new firm in a market of fixed size. Suppose that each firm can choose one of two different appearances for the product. The established producer prefers the newcomer's product to look different from its own (so that its customers will not be tempted to buy the newcomer's product), whereas the newcomer prefers that the products look alike. Or the game could model a relationship between two people in which one person wants to be like the other, whereas the other wants to be different.

? EXERCISE 18.1 (Games without conflict) Give some examples of two-player strategic games in which each player has two actions and the players have the same preferences, so that there is no conflict between their interests. (Present your games as tables like the one in Figure 17.1.)

2.5 Example: the *Stag Hunt*

A sentence in *Discourse on the origin and foundations of inequality among men* (1755) by the philosopher Jean-Jacques Rousseau discusses a group of hunters who wish to catch a stag. They will succeed if they all remain sufficiently attentive, but each is tempted to desert her post and catch a hare. One interpretation of the sentence is that the interaction between the hunters may be modeled as the following strategic game.

◆ EXAMPLE 18.2 (*Stag Hunt*) Each of a group of hunters has two options: she may remain attentive to the pursuit of a stag, or catch a hare. If all hunters pursue the stag, they catch it and share it equally; if any hunter devotes her energy to catching a hare, the stag escapes, and the hare belongs to the defecting hunter alone. Each hunter prefers a share of the stag to a hare.

The strategic game that corresponds to this specification is:

Players The hunters.

Actions Each player's set of actions is $\{Stag, Hare\}$.

Preferences For each player, the action profile in which all players choose *Stag* (resulting in her obtaining a share of the stag) is ranked highest, followed by any profile in which she chooses *Hare* (resulting in her obtaining a hare), followed by any profile in which she chooses *Stag* and one or more of the other players chooses *Hare* (resulting in her leaving empty-handed).

Like other games with many players, this game cannot easily be presented in a table like that in Figure 17.1. For the case in which there are two hunters, the game is shown in Figure 18.1.

	Stag	*Hare*
Stag	2, 2	0, 1
Hare	1, 0	1, 1

Figure 18.1 The *Stag Hunt* (Example 18.2) for the case of two hunters.

The variant of the two-player *Stag Hunt* shown in Figure 19.1 has been suggested as an alternative to the *Prisoner's Dilemma* as a model of an arms race, or, more generally, of the "security dilemma" faced by a pair of countries. The game differs from the *Prisoner's Dilemma* in that a country prefers the outcome in which both countries refrain from arming themselves to the one in which it alone arms itself: the cost of arming outweighs the benefit if the other country does not arm itself.

	Refrain	*Arm*
Refrain	3, 3	0, 2
Arm	2, 0	1, 1

Figure 19.1 A variant of the two-player *Stag Hunt* that models the "security dilemma".

2.6 Nash equilibrium

What actions will be chosen by the players in a strategic game? We wish to assume, as in the theory of a rational decision-maker (Section 1.2), that each player chooses the best available action. In a game, the best action for any given player depends, in general, on the other players' actions. So when choosing an action a player must have in mind the actions the other players will choose. That is, she must form a *belief* about the other players' actions.

On what basis can such a belief be formed? The assumption underlying the analysis in this chapter and the next two chapters is that each player's belief is derived from her past experience playing the game, and that this experience is sufficiently extensive that she *knows* how her opponents will behave. No one tells her the actions her opponents will choose, but her previous involvement in the game leads her to be sure of these actions. (The question of *how* a player's experience can lead her to the correct beliefs about the other players' actions is addressed briefly in Section 4.9.)

Although we assume that each player has experience playing the game, we assume that she views each play of the game in isolation. She does not become familiar with the behavior of specific opponents and consequently does not condition her action on the opponent she faces; nor does she expect her current action to affect the other players' future behavior.

It is helpful to think of the following idealized circumstances. For each player in the game there is a population of many decision-makers who may, on any occasion, take that player's role. In each play of the game, players are selected randomly, one from each population. Thus each player engages in the game repeatedly, against ever-varying opponents. Her experience leads her to beliefs about the actions of "typical" opponents, not any specific set of opponents.

As an example, think of the interaction between buyers and sellers. Buyers and sellers repeatedly interact, but to a first approximation many of the pairings may be modeled as random. In many cases a buyer transacts only once with any given seller, or interacts repeatedly but anonymously (when the seller is a large store, for example).

In summary, the solution theory we study has two components. First, each player chooses her action according to the model of rational choice, given her belief about the other players' actions. Second, every player's belief about the other players' actions is correct. These two components are embodied in the following definition.

> A *Nash equilibrium* is an action profile a^* with the property that no player i can do better by choosing an action different from a_i^*, given that every other player j adheres to a_j^*.

In the idealized setting in which the players in any given play of the game are drawn randomly from a collection of populations, a Nash equilibrium corresponds to a *steady state*. If, whenever the game is played, the action profile is the same Nash equilibrium a^*, then no player has a reason to choose any action different from her component of a^*; there is no pressure on the action profile to change. Expressed differently, a Nash equilibrium embodies a stable "social norm": if everyone else adheres to it, no individual wishes to deviate from it.

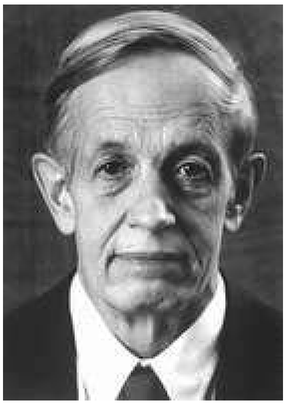
The second component of the theory of Nash equilibrium—that the players' beliefs about each other's actions are correct—implies, in particular, that two players' beliefs about a third player's action are the same. For this reason, the condition is sometimes said to be that the players' "expectations are coordinated".

The situations to which we wish to apply the theory of Nash equilibrium do not in general correspond exactly to the idealized setting described above. For example, in some cases the players do not have much experience with the game; in others they do not view each play of the game in isolation. Whether or not the notion of Nash equilibrium is appropriate in any given situation is a matter of judgment. In some cases, a poor fit with the idealized setting may be mitigated by other considerations. For example, inexperienced players may be able to draw conclusions about their opponents' likely actions from their experience in other situations, or from other sources. (One aspect of such reasoning is discussed in the box on page 30). Ultimately, the test of the appropriateness of the notion of Nash equilibrium is whether it gives us insights into the problem at hand.

With the aid of an additional piece of notation, we can state the definition of a Nash equilibrium precisely. Let a be an action profile, in which the action of each player i is a_i. Let a_i' be any action of player i (either equal to a_i, or different from it). Then (a_i', a_{-i}) denotes the action profile in which every player j *except* i chooses her action a_j as specified by a, whereas player i chooses a_i'. (The $-i$ subscript on a stands for "except i".) That is, (a_i', a_{-i}) is the action profile in which all the players other than i adhere to a while i "deviates" to a_i'. (If $a_i' = a_i$ then of course $(a_i', a_{-i}) = (a_i, a_{-i}) = a$.) If there are three players, for example, then (a_2', a_{-2}) is the action profile in which players 1 and 3 adhere to a (player 1 chooses a_1, player 3 chooses a_3) and player 2 deviates to a_2'.

Using this notation, we can restate the condition for an action profile a^* to be a Nash equilibrium: no player i has any action a_i for which she prefers (a_i, a_{-i}^*) to a^*. Equivalently, for every player i and every action a_i of player i, the action profile a^* is at least as good for player i as the action profile (a_i, a_{-i}^*).

JOHN F. NASH, JR.



A few of the ideas of John F. Nash Jr., developed while he was a graduate student at Princeton from 1948 to 1950, transformed game theory. Nash was born in 1928 in Bluefield, West Virginia, USA, where he grew up. He was an undergraduate mathematics major at Carnegie Institute of Technology from 1945 to 1948. In 1948 he obtained both a B.S. and an M.S., and began graduate work in the Department of Mathematics at Princeton University. (One of his letters of recommendation, from a professor at Carnegie Institute of Technology, was a single sentence: "This man is a genius" (Kuhn et al. 1995, 282).) A paper containing the main result of his thesis was submitted to the *Proceedings of the National Academy of Sciences* in November 1949, fourteen months after he started his graduate work. ("A fine goal to set … graduate students", to quote Harold Kuhn! (See Kuhn et al. 1995, 282.)) He completed his PhD the following year, graduating on his 22nd birthday. His thesis, 28 pages in length, introduces the equilibrium notion now known as "Nash equilibrium" and delineates a class of strategic games that have Nash equilibria (Proposition 117.1 in this book). The notion of Nash equilibrium vastly expanded the scope of game theory, which had previously focussed on two-player "strictly competitive" games (in which the players' interests are directly opposed). While a graduate student at Princeton, Nash also wrote the seminal paper in bargaining theory, Nash (1950b) (the ideas of which originated in an elective class in international economics he took as an undergraduate). He went on to take an academic position in the Department of Mathematics at MIT, where he produced "a remarkable series of papers" (Milnor 1995, 15); he has been described as "one of the most original mathematical minds of [the twentieth] century" (Kuhn 1996). He shared the 1994 Nobel prize in economics with the game theorists John C. Harsanyi and Reinhard Selten.

► DEFINITION 21.1 (*Nash equilibrium of strategic game with ordinal preferences*) The action profile a^* in a strategic game with ordinal preferences is a **Nash equilibrium** if, for every player i and every action a_i of player i, a^* is at least as good according to player i's preferences as the action profile (a_i, a^*_{-i}) in which player i chooses a_i while every other player j chooses a^*_j. Equivalently, for every player i,

$$u_i(a^*) \geq u_i(a_i, a^*_{-i}) \text{ for every action } a_i \text{ of player } i, \tag{21.2}$$

where u_i is a payoff function that represents player i's preferences.

This definition implies neither that a strategic game necessarily has a Nash equilibrium, nor that it has at most one. Examples in the next section show that some games have a single Nash equilibrium, some possess no Nash equilibrium, and others have many Nash equilibria.

The definition of a Nash equilibrium is designed to model a steady state among experienced players. An alternative approach to understanding players' actions in strategic games assumes that the players know each others' preferences, and considers what each player can deduce about the other players' actions from their rationality and their knowledge of each other's rationality. This approach is studied in Chapter 12. For many games, it leads to a conclusion different from that of Nash equilibrium. For games in which the conclusion is the same the approach offers us an alternative interpretation of a Nash equilibrium, as the outcome of rational calculations by players who do not necessarily have any experience playing the game.

STUDYING NASH EQUILIBRIUM EXPERIMENTALLY

The theory of strategic games lends itself to experimental study: arranging for subjects to play games and observing their choices is relatively straightforward. A few years after game theory was launched by von Neumann and Morgenstern's (1944) book, reports of laboratory experiments began to appear. Subsequently a huge number of experiments have been conducted, illuminating many issues relevant to the theory. I discuss selected experimental evidence throughout the book.

The theory of Nash equilibrium, as we have seen, has two components: the players act in accordance with the theory of rational choice, given their beliefs about the other players' actions, and these beliefs are correct. If every subject understands the game she is playing and faces incentives that correspond to the preferences of the player whose role she is taking, then a divergence between the observed outcome and a Nash equilibrium can be blamed on a failure of one or both of these two components. Experimental evidence has the potential of indicating the types of games for which the theory works well and, for those in which the theory does not work well, of pointing to the faulty component and giving us hints about the characteristics of a better theory. In designing an experiment that cleanly tests the theory, however, we need to confront several issues.

The model of rational choice takes preferences as given. Thus to test the theory of Nash equilibrium experimentally, we need to ensure that each subject's preferences are those of the player whose role she is taking in the game we are examining. The standard way of inducing the appropriate preferences is to pay each subject an amount of money directly related to the payoff given by a payoff function that represents the preferences of the player whose role the subject is taking. Such remuneration works if each subject likes money and cares only about the amount of money she receives, ignoring the amounts received by her opponents. The assumption that people like receiving money is reasonable in many cultures, but the assumption that people care only about their own monetary rewards—are "selfish"—may, in some contexts at least, not be reasonable. Unless we check whether our subjects are selfish in the context of our experiment, we will jointly test two hypotheses: that humans are selfish—a hypothesis not part of game theory—and that the notion of Nash equilibrium models their behavior. In some cases we may indeed wish to test these hypotheses jointly. But in order to test the theory of Nash equilibrium alone we need to ensure that we induce the preferences we wish to study.

Assuming that better decisions require more effort, we need also to ensure that each subject finds it worthwhile to put in the extra effort required to obtain a higher payoff. If we rely on monetary payments to provide incentives, the amount of money a subject can obtain must be sufficiently sensitive to the quality of her decisions to compensate her for the effort she expends (paying a flat fee, for example, is inappropriate). In some cases, monetary payments may not be necessary: under some circumstances, subjects drawn from a highly competitive culture like that of the USA may be sufficiently motivated by the possibility of obtaining a high score, even if that score does not translate into a monetary payoff.

The notion of Nash equilibrium models action profiles compatible with steady states. Thus to study the theory experimentally we need to collect observations of subjects' behavior when they have experience playing the game. But they should not have obtained that experience while knowingly facing the same opponents repeatedly, for the theory assumes that the players consider each play of the game in isolation, not as part of an ongoing relationship. One option is to have each subject play the game against many different opponents, gaining experience about how the other subjects on average play the game, but not about the choices of any other given player. Another option is to describe the game in terms that relate to a situation in which the subjects already have experience. A difficulty with this second approach is that the description we give may connote more than simply the payoff numbers of our game. If we describe the *Prisoner's Dilemma* in terms of cooperation on a joint project, for example, a subject may be biased toward choosing the action she has found appropriate when involved in joint projects, even if the structures of those interactions were significantly different from that of the *Prisoner's Dilemma*. As she plays the experimental game repeatedly she may come to appreciate how it differs from the games in which she has been involved previously, but her biases may disappear only slowly.

Whatever route we take to collect data on the choices of subjects experienced in playing the game, we confront a difficult issue: how do we know when the outcome has converged? Nash's theory concerns only equilibria; it has nothing to say about the path players' choices will take on the way to an equilibrium, and so gives us no guide as to whether 10, 100, or 1,000 plays of the game are enough to give a chance for the subjects' expectations to become coordinated.

Finally, we can expect the theory of Nash equilibrium to correspond to reality only approximately: like all useful theories, it definitely is not *exactly* correct. How do we tell whether the data are close enough to the theory to support it? One possibility is to compare the theory of Nash equilibrium with some other theory. But for many games there is no obvious alternative theory—and certainly not one with the generality of Nash equilibrium. Statistical tests can sometimes aid in deciding whether the data is consistent with the theory, though ultimately we remain the judge of whether or not our observations persuade us that the theory enhances our understanding of human behavior in the game.

2.7 Examples of Nash equilibrium

2.7.1 *Prisoner's Dilemma*

By examining the four possible pairs of actions in the *Prisoner's Dilemma* (reproduced in Figure 24.1), we see that (*Fink*, *Fink*) is the unique Nash equilibrium.

	Quiet	*Fink*
Quiet	2, 2	0, 3
Fink	3, 0	1, 1

Figure 24.1 The *Prisoner's Dilemma*.

The action pair (*Fink*, *Fink*) is a Nash equilibrium because (*i*) given that player 2 chooses *Fink*, player 1 is better off choosing *Fink* than *Quiet* (looking at the right column of the table we see that *Fink* yields player 1 a payoff of 1 whereas *Quiet* yields her a payoff of 0), and (*ii*) given that player 1 chooses *Fink*, player 2 is better off choosing *Fink* than *Quiet* (looking at the bottom row of the table we see that *Fink* yields player 2 a payoff of 1 whereas *Quiet* yields her a payoff of 0).

No other action profile is a Nash equilibrium:

- (*Quiet*, *Quiet*) does not satisfy (21.2) because when player 2 chooses *Quiet*, player 1's payoff to *Fink* exceeds her payoff to *Quiet* (look at the first components of the entries in the left column of the table). (Further, when player 1 chooses *Quiet*, player 2's payoff to *Fink* exceeds her payoff to *Quiet*: player 2, as well as player 1, wants to deviate. To show that a pair of actions is not a Nash equilibrium, however, it is not necessary to study player 2's decision once we have established that player 1 wants to deviate: it is enough to show that *one* player wishes to deviate to show that a pair of actions is not a Nash equilibrium.)

- (*Fink*, *Quiet*) does not satisfy (21.2) because when player 1 chooses *Fink*, player 2's payoff to *Fink* exceeds her payoff to *Quiet* (look at the second components of the entries in the bottom row of the table).

- (*Quiet*, *Fink*) does not satisfy (21.2) because when player 2 chooses *Fink*, player 1's payoff to *Fink* exceeds her payoff to *Quiet* (look at the first components of the entries in the right column of the table).

In summary, in the only Nash equilibrium of the *Prisoner's Dilemma* both players choose *Fink*. In particular, the incentive to free ride eliminates the possibility that the mutually desirable outcome $(Quiet, Quiet)$ occurs. In the other situations discussed in Section 2.2 that may be modeled as the *Prisoner's Dilemma*, the outcomes predicted by the notion of Nash equilibrium are thus as follows: both people goof off when working on a joint project; both duopolists charge a low price; both countries build bombs; both farmers graze their sheep a lot. (The overgrazing of a common thus predicted is sometimes called the "tragedy of the commons". The intuition that some of these dismal outcomes may be avoided if the same pair of people play the game repeatedly is explored in Chapter 14.)

In the *Prisoner's Dilemma*, the Nash equilibrium action of each player (*Fink*) is the best action for each player not only if the other player chooses her equilibrium action (*Fink*), but also if she chooses her other action (*Quiet*). The action pair $(Fink, Fink)$ is a Nash equilibrium because if a player believes that her opponent will choose *Fink* then it is optimal for her to choose *Fink*. But in fact it is optimal for a player to choose *Fink regardless* of the action she expects her opponent to choose. In most of the games we study, a player's Nash equilibrium action does not satisfy this condition: the action is optimal if the other players choose their Nash equilibrium actions, but some other action is optimal if the other players choose non-equilibrium actions.

? EXERCISE 25.1 (Variant of *Prisoner's Dilemma* with altruistic preferences) Each of two players has two possible actions, *Quiet* and *Fink*; each action pair results in the players' receiving amounts of *money* equal to the numbers corresponding to that action pair in Figure 24.1. (For example, if player 1 chooses *Quiet* and player 2 chooses *Fink*, then player 1 receives nothing, whereas player 2 receives $3.) The players are not "selfish"; rather, the preferences of each player i are represented by the payoff function $m_i(a) + \alpha m_j(a)$, where $m_i(a)$ is the amount of money received by player i when the action profile is a, j is the other player, and α is a given nonnegative number. Player 1's payoff to the action pair $(Quiet, Quiet)$, for example, is $2 + 2\alpha$.

a. Formulate a strategic game that models this situation in the case $\alpha = 1$. Is this game the *Prisoner's Dilemma*?

b. Find the range of values of α for which the resulting game is the *Prisoner's Dilemma*. For values of α for which the game is not the *Prisoner's Dilemma*, find its Nash equilibria.

? EXERCISE 25.2 (Selfish and altruistic social behavior) Two people enter a bus. Two adjacent cramped seats are free. Each person must decide whether to sit or stand. Sitting alone is more comfortable than sitting next to the other person, which is more comfortable than standing.

a. Suppose that each person cares only about her own comfort. Model the situation as a strategic game. Is this game the *Prisoner's Dilemma*? Find its Nash equilibrium (equilibria?).

b. Suppose that each person is altruistic, ranking the outcomes according to the *other* person's comfort, but, out of politeness, prefers to stand than to sit if the other person stands. Model the situation as a strategic game. Is this game the *Prisoner's Dilemma*? Find its Nash equilibrium (equilibria?).

c. Compare the people's comfort in the equilibria of the two games.

EXPERIMENTAL EVIDENCE ON THE *Prisoner's Dilemma*

The *Prisoner's Dilemma* has attracted a great deal of attention by economists, psychologists, sociologists, and biologists. A huge number of experiments have been conducted with the aim of discovering how people behave when playing the game. Almost all these experiments involve each subject's playing the game repeatedly against an unchanging opponent, a situation that calls for an analysis significantly different from the one in this chapter (see Chapter 14).

The evidence on the outcome of isolated plays of the game is inconclusive. No experiment of which I am aware carefully induces the appropriate preferences and is specifically designed to elicit a steady state action profile (see the box on page 22). Thus in each case the choice of *Quiet* by a player could indicate that she is not "selfish" or that she is not experienced in playing the game, rather than providing evidence against the notion of Nash equilibrium.

In two experiments with very low payoffs, each subject played the game a small number of times against different opponents; between 50% and 94% of subjects chose *Fink*, depending on the relative sizes of the payoffs and some details of the design (Rapoport, Guyer, and Gordon 1976, 135–137, 211–213, and 223-226). In a more recent experiment, 78% of subjects chose *Fink* in the last 10 of 20 rounds of play against different opponents (Cooper, DeJong, Forsythe, and Ross 1996). In face-to-face games in which communication is allowed, the incidence of the choice of *Fink* tends to be lower: from 29% to 70% depending on the nature of the communication allowed (Deutsch 1958, and Frank, Gilovich, and Regan 1993, 163–167). (In all these experiments, the subjects were college students in the USA or Canada.)

One source of the variation in the results seems to be that some designs induce preferences that differ from those of the *Prisoner's Dilemma*; no clear answer emerges to the question of whether the notion of Nash equilibrium is relevant to the *Prisoner's Dilemma*. If, nevertheless, one interprets the evidence as showing that some subjects in the *Prisoner's Dilemma* systematically choose *Quiet* rather than *Fink*, one must fault the rational choice component of Nash equilibrium, not the coordinated expectations component. Why? Because, as noted in the text, *Fink* is optimal *no matter* what a player thinks her opponent will choose, so that any model in which the players act according to the model of rational choice, whether or not their expectations are coordinated, predicts that each player chooses *Fink*.

2.7.2 BoS

To find the Nash equilibria of *BoS* (Figure 16.1), we can examine each pair of actions in turn:

- (*Bach*, *Bach*): If player 1 switches to *Stravinsky* then her payoff decreases from 2 to 0; if player 2 switches to *Stravinsky* then her payoff decreases from 1 to 0. Thus a deviation by either player decreases her payoff. Thus (*Bach*, *Bach*) is a Nash equilibrium.

- (*Bach*, *Stravinsky*): If player 1 switches to *Stravinsky* then her payoff increases from 0 to 1. Thus (*Bach*, *Stravinsky*) is not a Nash equilibrium. (Player 2 can increase her payoff by deviating, too, but to show the pair is not a Nash equilibrium it suffices to show that one player can increase her payoff by deviating.)

- (*Stravinsky*, *Bach*): If player 1 switches to *Bach* then her payoff increases from 0 to 2. Thus (*Stravinsky*, *Bach*) is not a Nash equilibrium.

- (*Stravinsky*, *Stravinsky*): If player 1 switches to *Bach* then her payoff decreases from 1 to 0; if player 2 switches to *Bach* then her payoff decreases from 2 to 0. Thus a deviation by either player decreases her payoff. Thus (*Stravinsky*, *Stravinsky*) is a Nash equilibrium.

We conclude that the game has two Nash equilibria: (*Bach*, *Bach*) and (*Stravinsky*, *Stravinsky*). That is, both of these outcomes are compatible with a steady state; both outcomes are stable social norms. If, in every encounter, both players choose *Bach*, then no player has an incentive to deviate; if, in every encounter, both players choose *Stravinsky*, then no player has an incentive to deviate. If we use the game to model the choices of men when matched with women, for example, then the notion of Nash equilibrium shows that two social norms are stable: both players choose the action associated with the outcome preferred by women, and both players choose the action associated with the outcome preferred by men.

2.7.3 Matching Pennies

By checking each of the four pairs of actions in *Matching Pennies* (Figure 17.1) we see that the game has no Nash equilibrium. For the pairs of actions (*Head*, *Head*) and (*Tail*, *Tail*), player 2 is better off deviating; for the pairs of actions (*Head*, *Tail*) and (*Tail*, *Head*), player 1 is better off deviating. Thus for this game the notion of Nash equilibrium isolates no steady state. In Chapter 4 we return to this game; an extension of the notion of a Nash equilibrium gives us an understanding of the likely outcome.

2.7.4 The Stag Hunt

Inspection of Figure 18.1 shows that the two-player *Stag Hunt* has two Nash equilibria: (*Stag*, *Stag*) and (*Hare*, *Hare*). If one player remains attentive to the pursuit of the stag, then the other player prefers to remain attentive; if one player chases a hare, the other one prefers to chase a hare (she cannot catch a stag alone). (The equilibria of the variant of the game in Figure 19.1 are analogous: (*Refrain*, *Refrain*) and (*Arm*, *Arm*).)

Unlike the Nash equilibria of *BoS*, one of these equilibria is better for both players than the other: each player prefers (*Stag*, *Stag*) to (*Hare*, *Hare*). This fact has no bearing on the equilibrium status of (*Hare*, *Hare*), since the condition for an equilibrium is that a *single* player cannot gain by deviating, *given* the other player's behavior. Put differently, an equilibrium is immune to any *unilateral* deviation; coordinated deviations by groups of players are not contemplated. However, the existence of two equilibria raises the possibility that one equilibrium might more likely be the outcome of the game than the other. I return to this issue in Section 2.7.6.

I argue that the many-player *Stag Hunt* (Example 18.2) also has two Nash equilibria: the action profile (*Stag*, . . . , *Stag*) in which every player joins in the pursuit of the stag, and the profile (*Hare*, . . . , *Hare*) in which every player catches a hare.

- (*Stag*, . . . , *Stag*) is a Nash equilibrium because each player prefers this profile to that in which she alone chooses *Hare*. (A player is better off remaining attentive to the pursuit of the stag than running after a hare if all the other players remain attentive.)
- (*Hare*, . . . , *Hare*) is a Nash equilibrium because each player prefers this profile to that in which she alone pursues the stag. (A player is better off catching a hare than pursuing the stag if no one else pursues the stag.)
- No other profile is a Nash equilibrium, because in any other profile at least one player chooses *Stag* and at least one player chooses *Hare*, so that any player choosing *Stag* is better off switching to *Hare*. (A player is better off catching a hare than pursing the stag if at least one other person chases a hare, since the stag can be caught only if everyone pursues it.)

? EXERCISE 28.1 (Variants of the *Stag Hunt*) Consider two variants of the n-hunter *Stag Hunt* in which only m hunters, with $2 \leq m < n$, need to pursue the stag in order to catch it. (Continue to assume that there is a single stag.) Assume that a captured stag is shared only by the hunters who catch it.

a. Assume, as before, that each hunter prefers the fraction $1/n$ of the stag to a hare. Find the Nash equilibria of the strategic game that models this situation.

b. Assume that each hunter prefers the fraction $1/k$ of the stag to a hare, but prefers the hare to any smaller fraction of the stag, where k is an integer with $m \leq k \leq n$. Find the Nash equilibria of the strategic game that models this situation.

The following more difficult exercise enriches the hunters' choices in the *Stag Hunt*. This extended game has been proposed as a model that captures Keynes' basic insight about the possibility of multiple economic equilibria, some of which are undesirable (Bryant 1983, 1994).