Game Theory for Political Scientists

James D. Morrow

Princeton University Press Princeton, New Jersey

useful. Black 1958 is the source that began the formal study of committee decisions.

The structure-induced equilibrium model has generated a large literature. Historically, the notable papers are Shepsle 1979, Shepsle and Weingast 1987, and Weingast 1989. Hammond and Miller (1987) use this approach to analyze the organization of Congress given in the U.S constitution.

The model of bargaining in legislatures by Baron and Ferejohn (1989) has led to a set of papers analyzing the distributive effects of congressional policies (Baron 1989a, 1991b).

The third model of congressional structure is informational; it is presented in Chapter Eight. Krehbiel 1991 is the place to begin; Gilligan and Krehbiel (1987, 1989, 1990) present their models of an informative committee structure. Huber (1992) uses models of congressional rules to explain differences between the legislatures of France and the United States.

The other institutions of Congress have also been modeled. Calvert (1987) uses the Chain Store Paradox (see Chapter Nine) to analyze legislative leadership. McKelvey and Riezman (1992) explain why legislators use the seniority system to distribute committee positions. Austen-Smith (1990) and Austen-Smith and Riker (1987, 1990) present models of legislative debate based on signaling theory. Sullivan (1990) presents a simple model of bargaining between the president and the Congress. The section on models of bureaucracy and administration in Chapter Nine of this book includes models of legislative-agency relations. These models analyze congressional oversight and control of agencies.

Ainsworth and Sened (1993) and Austen-Smith (1993) present models of lobbying as an information source for legislators. Austen-Smith (1992) considers how legislative constituencies could restrict sophisticated votes in both chambers of congress.

Chapter Six
Beliefs and Perfect Bayesian Equilibria

Perfect Bayesian equilibrium unites a new concept, beliefs, with strategies to create a more powerful idea of equilibrium. So far, equilibria have been combinations of strategies that are best replies to one another. Chapter Five added the idea that best replies should be judged off the equilibrium path as well as on it. Backwards induction and subgame perfection provided one way to judge best replies off the equilibrium path. But information sets with multiple nodes often frustrate those techniques. Because we cannot do a backwards induction through information sets with multiple nodes, backwards induction is powerless to deal with many games. Subgame perfection can help in some of these cases but not all.

Information sets with multiple nodes reflect a player's inability to verify the node it is located at when it must choose. It cannot determine the consequences of its moves because of this uncertainty. When a player chooses at an information set with multiple nodes, it uses what information it has about which node it is likely to be at when it chooses. We represent these judgments about what has happened with conditional probabilities. A player thinks to itself, *What is the probability that I am at the top node in this information set given that I must make a move from this information set?*

These conditional probabilities on the nodes of an information set with multiple nodes are called **beliefs**. They summarize a player's judgment about what has probably happened up to that point in the game. Beliefs express the likelihood that the moving player is at each node in such an information set. We use beliefs to calculate a player's expected utility for each action from an information set with multiple nodes. For each action, we weigh the utility of the outcome that results from that action from each node in the information set by the probability the player is at that node.

Not just any probabilities can be considered rational beliefs in the context of a particular equilibrium. Beliefs should reflect what the players know about the game and the common conjecture they hold about the strategies they are playing. They capture the moving player's hypotheses about the history of play that led to this information set. A player's beliefs reproduce what it believes the other players are likely to have done before it must move.

Further, we assume that players make optimal use of the information available to them. We model how players use information to revise their beliefs with

Bayes's Theorem from probability theory. When actors are uncertain, they use new information to update their beliefs about underlying states of the world. Rarely is new information decisive for judging the state of the world. Instead, it shifts the player's judgment about the likelihood of different states. Bayes's Theorem explains how new information should be used to update beliefs about underlying states of the world. It weighs the ability of the new evidence to discriminate among different states and the strength of prior beliefs to arrive at updated beliefs.

The states in a game are all the moves that have been made earlier in the game. If a player knew what moves had been made, it could determine which node it is at when it must move. When it cannot verify what moves have been made, it uses the other players' equilibrium strategies to judge at which node it is likely to be. A player updates its beliefs by combining what it can observe with the likelihood in the equilibrium that the other players' would make each possible move. Bayes's Theorem is the formal tool for this updating. Rational beliefs, then, depend on the players' strategies. Along the equilibrium path, the beliefs should be calculated from the known probabilities of chance moves and the players' strategies by means of Bayes's Theorem.

A perfect Bayesian equilibrium consists of beliefs and strategies that support one another in equilibrium. Given the beliefs and the other players' strategies, each player's strategy is optimal at every node in the game and all beliefs are consistent with the equilibrium strategies along the equilibrium path.

The addition of beliefs reinforces the tie between game theory and decision theory. Expected utility calculations model decisions in the sequence of the game. Beliefs add a concept that seems very natural for analyzing games. We can analyze how beliefs and actions are related in equilibrium. We can trace how actions change beliefs and how beliefs lead to actions in a perfect Bayesian equilibrium. Beliefs provide an intuitive way to discuss how players deal with incomplete information and incorporate a form of learning into a game.

The difficulty of the material steps up here. Like most of game theory, the mathematics are not difficult, generally requiring just algebra. But careful attention to the mathematics is required here even though the ideas are quite simple and intuitive. Translating those ideas into a careful formal argument requires close attention to the mathematical details. Calculations of both expected utility and Bayesian updating are necessary to find equilibria. These calculations are not difficult, but there is no substitute for them. Further, strategic logic can be complex. Figuring out the equation that expresses the correct strategic calculation is often more difficult than solving that equation. Formal solutions are necessary here precisely because intuition alone is often wrong. The discipline of formalization is needed to structure and shape our intuition. Working through these models compels us to think carefully about the incentives the players face and what strategies can best achieve their goals.

This chapter begins by reviewing Bayes's Theorem. Bayes's Theorem is essential for updating beliefs in a perfect Bayesian equilibrium. I follow Bayes's Theorem with an example of Bayesian decision theory, the preference for biased information. Bayesian decision theory links expected utility calculations and Bayesian updating of beliefs to see how new information can change decisions. I then introduce beliefs with a classic example from Selten 1975. I define perfect Bayesian equilibrium next. The chapter ends with an example drawn from nuclear deterrence.

Bayes's Theorem*

Deciders are often uncertain about the consequences of their actions. We represent their uncertainty by subjective probabilities over the possible states of the world. These probabilities represent a decider's degree of belief about the likelihood of each different state. The higher the subjective probability of a state, the more likely is the decider to believe that state is the true state of the world.

These beliefs should change as a decider gains new information about the state of the world. Sometimes, that information may convince the decider about the true state of the world. If event E can happen only when A is the state of the world, then observing E is sufficient to conclude that A is the state of the world. But new information rarely allows such strong conclusions. Typically, an event could occur under several different states of the world with differing probabilities. Actors use the probabilities of an observed event's occurring given each possible state to update their probabilities about the state of the world. The subjective probabilities of each state before consideration of new information are called **prior (or initial) beliefs (or probabilities)**. Updating considers both the prior beliefs and the probabilities that the event will occur given each state. The updated beliefs are called **posterior beliefs (or probabilities)**. Some events are more likely under some states of the world than under others. Observing an event provides information that increases beliefs about states of the world where it is more likely to occur.

Beliefs are conditional probabilities. Let A be a state of the world and B an event. The probability of B given A, written $p(B|A)$, specifies the likelihood that B will occur given that A is the state of the world. Bayes's Theorem uses the conditional probabilities of events given states to deduce the conditional probabilities of states given events. For instance, the latter would be the conditional probability of A given B, written $p(A|B)$.

*This section uses conditional probabilities.

Theorem (Bayes's Theorem): Let $(A_i)_{i=1}^n$ be the set of states of the world, and B an event. Then

$$p(A_i|B) = \frac{p(A_i)p(B|A_i)}{\sum_{i=1}^{n} p(A_i)p(B|A_i)}$$

If there are only two states of the world, A and not A (abbreviated ~ A), the above formula simplifies to

$$p(A|B) = \frac{p(A)p(B|A)}{p(A)p(B|A) + p(\sim A)p(B|\sim A)}$$

Bayes's Theorem determines the posterior probability of a state by calculating the probability that both the event and the state will occur and dividing it by the probability that the event will occur regardless of state (determined by summing across all states). An event, B in the formula, changes beliefs about the underlying state of the world, A in the formula, because different states produce different probabilities of the event's occurring. We learn about the state of the world by observing events that are more likely to occur under one state than under others. If an event is equally likely under all states, prior beliefs will not change after observing the event. Events with greater differences in probability given each state discriminate more effectively across the states than events with small differences.

Bayes's Theorem follows directly from the definition of a conditional probability. The probability of A given B, written p(A|B), is the probability of (A and B) divided by the probability of B, p(A and B)/p(B). The probability of (A and B) is p(A)p(B|A), from the definition of the conditional probability of B given A, p(B|A) = p(A and B)/p(A). The probability of B is the sum of the probabilities that (A and B) occur and that [(not A) and B] occur. These probabilities are p(A)p(B|A) and p(~A)p(B|~A), respectively. Substituting these probabilities into the conditional probability of A given B gives us Bayes's Theorem.

Example: Commissioner Crackdown wants to rid baseball of all players using drugs through drug testing. A particular test detects drug use successfully 90 percent of the time, but gives a false positive (i.e., a player tests positive even though that player has not used drugs) 10 percent of the time. If 10 percent of all players use drugs, what is the probability that a randomly selected player who tests positive is using drugs?

Let D signify that a player uses drugs, ~D signify that a player does not use drugs, and + signify that a player had a positive test result. We want to know p(D|+), the probability that a player uses drugs given that that player has tested positive.

$$p(D|+) = \frac{p(D)p(+|D)}{p(D)p(+|D) + p(\sim D)p(+|\sim D)} = \frac{(.1)(.9)}{(.1)(.9) + (.9)(.1)} = .5.$$

There is a 50 percent chance that a player testing positive has used drugs.

Example: In a certain city, 30 percent of the people are conservatives, 50 percent are liberals, and 20 percent are independents. Records show that in the latest election, 65 percent of the conservatives, 82 percent of the liberals, and 50 percent of the independents voted. If a person in the city is selected at random and it is learned that he or she did not vote in the last election, what is the probability that he or she is a liberal?

We want to know p(L|~v), where L signifies that the voter is a liberal and ~v signifies that he or she did not vote.

$$p(L|\sim v) = \frac{p(L)p(\sim v|L)}{p(C)p(\sim v|C) + p(L)p(\sim v|L) + p(I)p(\sim v|I)}$$
$$= \frac{(.5)(.18)}{(.3)(.35) + (.5)(.18) + (.2)(.5)} = \frac{18}{59}.$$

Exercise 6.1: A bag contains a thousand coins. One of the coins is badly loaded, so that it comes up heads $\frac{3}{4}$ of the time. A coin is drawn at random. What is the probability that it is the loaded coin if it is flipped and turns up heads without fail

a) three times in a row?

b) ten times in a row?

c) twenty times in a row?

In game theory, the states are the other players' strategies, and the events are the moves observed. If one player knows another player's strategy, it can predict all of the other player's future moves (up to any randomization through mixed strategies). A player's moves can reveal its strategy to the other players. Mixed strategies lead to partial, rather than total, revelation of strategy. Other players can use the information in the observed moves to infer the

strategy of the first players. Those other players can then adjust their own strategies in response. Bayes's Theorem is the formal tool used to model this updating in a game. Each player has an initial probability distribution over the other players' pure strategies that reflects its beliefs about what they will do. Each strategy specifies a probability for each action at each node (pure strategies give probabilities of 0 and 1 for each action). After observing a move by another player, a player uses its prior beliefs, the set of possible strategies, and Bayes's Theorem to calculate new probabilities for each strategy of the moving player.

The Preference for Biased Information

Bayesian decision theory gives us a way to explore how information affects choices. Does new information change a decider's choice from the one it would make without that information? Bayes's Theorem allows us to update the decider's subjective probability distribution and thus determine if its decision changes. I explore a related question in this section, the choice among different sources of information. Given a choice among possible sources of information, which source is most likely to affect a decision? If consulting sources of information is costly, which sources should be consulted? The best source is the one that is most likely to shift one's decision from what would be chosen in the absence of new information. The following model is a simplification of Calvert 1985.

Consider the position of a decider choosing between two courses of action, A_1 and A_2, in the face of uncertainty about the desirability of each course of action. The actual desirability of each course of action, denoted by x_1 and x_2, respectively, is either 0 or 1. The decider's payoff is the desirability of the chosen action: x_1 if A_1 is chosen or x_2 if A_2 is chosen. The actual values of x_1 and x_2 are not observed by the decider. Instead, it can consult an advisor who produces a recommendation about the desirability of each course of action based on the actual desirability of each.

The decider believes x_1 is better before receiving any advice. We represent this in the decider's prior beliefs. It believes that A_1 is more likely to produce a desirable outcome than A_2. This bias is best thought of as the decider's existing belief that A_1 is more effective than A_2. This bias may be an understanding that A_1 is generally a better option than A_2. Extensive experience with both options in prior settings could create such an understanding. It should not be thought of as a blind prejudice of the decider. The decider's initial beliefs are as follows:

$$p(x_1 = 1) = \tfrac{2}{3} \qquad p(x_1 = 0) = \tfrac{1}{3}$$

$$p(x_2 = 1) = \tfrac{1}{3} \qquad p(x_2 = 0) = \tfrac{2}{3}$$

An advisor can provide a "good" or "bad" recommendation for each alternative after observing its true desirability. Advisors are not strategic actors. An advisor produces recommendations based on the desirability of an action and a built-in bias, α, it has in favor of A_1 and against A_2. Any advisor makes some errors in its recommendations. Advisors sometimes say that an option is "bad" when $x_i = 1$ and that it is "good" when $x_i = 0$ for $i = 1, 2$. Formally, we have the following probabilities for recommendations:

$$p(A_1 \text{ good}|x_1 = 1) = \left(\tfrac{2}{3}\right)^{\frac{1}{\alpha}} \qquad p(A_1 \text{ good}|x_1 = 0) = \left(\tfrac{1}{3}\right)^{\frac{1}{\alpha}}$$

$$p(A_2 \text{ good}|x_2 = 1) = \left(\tfrac{2}{3}\right)^{\alpha} \qquad p(A_2 \text{ good}|x_2 = 0) = \left(\tfrac{1}{3}\right)^{\alpha}$$

The probabilities of bad recommendations are 1 − (probability of a good recommendation).

The parameter α gives an advisor's bias in favor of A_1 and against A_2. If $\alpha = 1$, the advisor gives neutral recommendations. The probability of a "good" recommendation by an unbiased advisor is $\frac{2}{3}$ and the probability of a "bad" recommendation is $\frac{1}{3}$. As $\alpha > 1$ increases, the advisor is more likely to say that A_1 is "good" regardless of the true value of x_1 and less likely to say that A_2 is "good." However, biased advisors are honest in the sense that they are more likely to say an option A_i is "good" when $x_i = 1$ than when $x_i = 0$ for both options.

We want to know what action the decider selects after receiving advice from the different advisors. If collecting advice is costly, then the decider benefits from advice only when that advice convinces it to change its decision. We compare the advice from two possible advisors, one unbiased and one biased in favor of A_1. For each possible piece of advice, we calculate the decider's updated distribution of the efficacy of each action. It chooses the action with the higher expected outcome.

Consider the unbiased advisor first. If the unbiased advisor recommends that A_1 is "good," we calculate the decider's posterior probabilities for x_1. Bayes's Theorem is used as follows to calculate these posterior probabilities:

$$\begin{aligned} &p(x_1 = 1|A_1 \text{ good}) \\ &= \frac{p(x_1 = 1)\,p(A_1 \text{ good}|x_1 = 1)}{p(x_1 = 1)\,p(A_1 \text{ good}|x_1 = 1) + p(x_1 = 0)\,p(A_1 \text{ good}|x_1 = 0)} \\ &= \frac{(\frac{2}{3})(\frac{2}{3})}{(\frac{2}{3})(\frac{2}{3}) + (\frac{1}{3})(\frac{1}{3})} = \frac{4}{5}. \end{aligned}$$

The probability that $x_1 = 0$ given that the unbiased advisor recommends that A_1 is "good" is 1 − (the probability above) = $\frac{1}{5}$.

With the posterior probability distribution, we calculate the decider's expectation for choosing A_1 after receiving a "good" recommendation. We sum

the value of each possible outcome by the probability of its occurring after the decider receives the unbiased advisor's recommendation that A_1 is "good":

$$E(A_1|A_1 \text{ good}) = p(x_1 = 1|A_1 \text{ good})x_1 + p(x_1 = 0|A_1 \text{ good})x_1$$
$$= \left(\frac{4}{5}\right)(1) + \left(\frac{1}{5}\right)(0) = \frac{4}{5}.$$

We calculate the decider's expectation for choosing each action after receiving each possible recommendation about that action in parallel fashion. Calculate the posterior distribution after receiving each recommendation, and use those probabilities to calculate an expected value. These three expectations are as follows:

$$E(A_1|A_1 \text{ bad}) = \tfrac{1}{2} \quad E(A_2|A_2 \text{ good}) = \tfrac{1}{2} \quad E(A_2|A_2 \text{ bad}) = \tfrac{1}{5}$$

Exercise 6.2: Verify that each of the three expectations above is correct.

Because the decider can choose only one course of action, it always chooses A_1. Its expected utility for choosing A_1 is always at least as great as that for choosing A_2, even if the neutral advisor advises that A_1 is "bad" and A_2 "good." If advice is costly, the decider should never consult the neutral advisor. Advice from the neutral advisor never leads the decider to change its chosen action from its prior belief. Why pay for advice that makes no difference?

But what about the biased advisor? Let $\alpha = 2$. Once again, we calculate the decider's expectation for each course of action after receiving each type of recommendation from the biased advisor. For a "good" recommendation for A_1, we have the following calculation for the decider's belief about the efficacy of A_1:

$$p(x_1 = 1|A_1 \text{ good})$$
$$= \frac{p(x_1 = 1)p(A_1 \text{ good}|x_1 = 1)}{p(x_1 = 1)p(A_1 \text{ good}|x_1 = 1) + p(x_1 = 0)p(A_1 \text{ good}|x_1 = 0)}$$
$$= \frac{(\frac{2}{3})(\frac{2}{3})^{\frac{1}{2}}}{(\frac{2}{3})(\frac{2}{3})^{\frac{1}{2}} + (\frac{1}{3})(\frac{1}{3})^{\frac{1}{2}}} = \frac{2\sqrt{2}}{2\sqrt{2}+1} \approx .74.$$

With this probability, we can calculate the decider's expected utility for choosing A_1 after it receives a "good" recommendation:

$$E(A_1|A_1 \text{ good}) = p(x_1 = 1|A_1 \text{ good})x_1 + p(x_1 = 0|A_1 \text{ good})x_1$$
$$= \left(\frac{2\sqrt{2}}{2\sqrt{2}+1}\right)(1) + \left(\frac{1}{2\sqrt{2}+1}\right)(0) = \frac{2\sqrt{2}}{2\sqrt{2}+1} \approx .74.$$

For a good recommendation for A_2, the calculation is similar, as follows:

$$p(x_2 = 1|A_2 \text{ good})$$
$$= \frac{p(x_2 = 1)p(A_2 \text{ good}|x_2 = 1)}{p(x_2 = 1)p(A_2 \text{ good}|x_2 = 1) + p(x_2 = 0)p(A_2 \text{ good}|x_2 = 0)}$$
$$= \frac{(\frac{1}{3})(\frac{2}{3})^2}{(\frac{1}{3})(\frac{2}{3})^2 + (\frac{2}{3})(\frac{1}{3})^2} = \frac{2}{3},$$

so

$$E(A_2|A_2 \text{ good}) = p(x_2 = 1|A_2 \text{ good})x_2 + p(x_2 = 0|A_2 \text{ bad})x_2$$
$$= (\tfrac{2}{3})(1) + (\tfrac{1}{3})(0) = \tfrac{2}{3}.$$

The following results are found by carrying out the calculations for the remaining two cases:

$$E(A_1|A_1 \text{ bad}) = \frac{2(\sqrt{3}-\sqrt{2})}{3\sqrt{3}-2\sqrt{2}-1} \approx .46,$$

and

$$E(A_2|A_2 \text{ bad}) = \tfrac{5}{21}.$$

Exercise 6.3: Verify that each of the two expectations above is correct.

The biased advisor can produce decisive advice. The decider will choose A_2 if the biased advisor says A_1 is "bad" and A_2 is "good". The biased advice may be worth paying for (depending on its price). This result may seem strange—the best advisors may be those who share the same biases as the decider. The practical advice is to surround yourself with advisors who share your biases but still retain some integrity. The intuition behind this result is that people discount different sources of information when they know the biases of those sources. The biased source is more useful than the neutral source because it is unlikely to say that A_1 is "bad" and A_2 is "good." When it does, the decider's beliefs about the value of both options shift dramatically. That

recommendation is sufficient to overwhelm the initial bias of the decider in favor of A_1 and cause it to choose A_2. Because the biased advisor rarely produces such a recommendation, that recommendation carries much weight in the eyes of the decider. The neutral source, in a sense, sends too many signals. The decider discounts its recommendations for A_2 and against A_1 because such signals are common. Those recommendations from the neutral source are sufficiently frequent that they fail to convince the decider that A_1 is a bad option compared to A_2. The amount of information such signals convey is insufficient to overcome the decider's existing bias in favor of A_1.

This result is particularly interesting because it goes against common sense. Psychological studies show that individuals often rely on sources of information that share the individual's biases–behavior referred to as "bolstering." Some have argued that bolstering is evidence that individuals are irrational because "rational" actors should look for neutral sources of information. This model suggests that the rational selection of information sources is not so simple. Biased sources may often be the best sources because advice against their biases is a clear signal to change actions. During the Vietnam War, it was no surprise to President Johnson that Senator Fulbright was opposed to the war. Consequently, Fulbright's opposition to the war carried little weight with Johnson. But when Robert McNamara came out against the war in 1967, the change in position by an original "hawk" in favor of the war had a strong effect on Johnson's evaluation of the war. Of course, this observation requires that the biased source must retain some honesty. A flunky who always provides an optimistic review of the options is useless.

It is not clear how general is the preference for biased information. This result depends upon the specific assumptions of this model. Changing some of the details of the model makes the unbiased source preferable. However, the intuition does seem general. Consider sources with a bias opposite from the decider's preexisting judgment. When you know a source of information is opposed to your own inclination, you expect the source will produce recommendations against your bias. One should rationally discount recommendations from such a source; it is biased. Any unusual recommendation from that source merely reinforces your confidence in your existing bias. Individuals may be quite rational when they select sources of information that share their own biases. Only those sources can produce evidence that will convince them to change their position on the options.

Perfect Bayesian Equilibria

Subgame perfection forces players to be rational in every subgame. But not every move begins a proper subgame, and subgame perfection cannot judge the rationality of behavior at such moves. For example, no proper subgame

can begin at an information set that includes more than one node. A player could make a noncredible threat at that information set and use that threat to deter the other player at a preceding node. How can we judge whether moves at such information sets are rational?

Perfect Bayesian equilibrium resolves this problem by introducing the concept of beliefs. When a player reaches a singleton information set, it knows the entire history of the game to that point. It decides which move is optimal by using the other players' strategies to predict their future moves, and thus predict the outcome of each possible move. It calculates its expected utility for each available move to choose its move. When a player reaches an information set with multiple nodes, its optimal move often varies with the node reached. A move may be optimal from one node but not from another. We cannot be certain which move is optimal from that information set because we do not know at which node the player is.

Beliefs solve this problem by allowing us to weigh the different nodes in an information set, and then calculate the player's expected utility from that information set. A player's beliefs are represented by a probability distribution over the nodes in an information set. For a given information set, they specify the probability that the player is at each node if the information set is reached. The player's expected utility for each available move is calculated by using these probabilities. We weigh the expected utility of each available action from each node in the information set by the player's belief that it is at that node, and then sum across all nodes in the information set. A player chooses the action that maximizes its expected utility.

Beliefs for an information set capture the players' hypotheses about the current state of the game. Beliefs are required to be consistent with equilibrium strategies wherever possible. On the equilibrium path, beliefs are the probabilities each node will be reached in the equilibrium. Off the equilibrium path, beliefs reflect hypotheses about what defections from the equilibrium led to those nodes. Beliefs reflect judgments about both the outcomes of prior, but still secret, chance moves and prior, yet unknown, strategy choices of the other players. The players use one another's strategies to predict the consequences of their own moves in any form of equilibrium. A player's judgments about the other players' strategies are captured in its beliefs and moves. Perfect Bayesian equilibria, then, create a symbiotic relationship between strategies and beliefs; in equilibrium, strategies are optimal given the beliefs, and the beliefs are consistent with the strategies.

Before formalizing this notion of equilibrium, I present an example of how beliefs can address the rationality of moves in games with information sets with multiple nodes.

Example: Consider the game in Figure 6.1 from Selten 1975. One Nash equilibrium of this game is (D;a;*L*). Each player's move is a

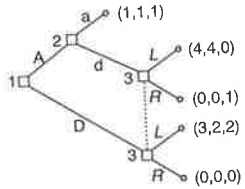

Figure 6.1 Selten's Game

best reply on the equilibrium path. If Player 2 will play a and Player 3 will play L, then Player 1 prefers D to A. D produces a payoff of 3 for him, whereas A gives him a payoff of 1. If Player 1 plays D, Player 2's move is off the equilibrium path. In a Nash equilibrium, any move off the equilibrium path is a best reply. Player 3's move is optimal for her if Player 1 plays D. Then Player 3 will have to move, and she will be at her lower node. L produces a payoff of 2 for Player 3, while R produces 0 for her.

Consider Player 2's position in this Nash equilibrium. If Player 2 has to make a move, then Player 1 must have played A. Player 2's move should be based solely on her expectation about Player 3's move. If Player 3 plays L with probability greater than $\frac{1}{4}$, Player 2 should play d instead of a. But (D;d;L) is not a Nash equilibrium. Player 1 wants to change his strategy from D to A, leading Player 3 to change from L to R, and so on. However, Player 2 never plays a in (D;a;L), so playing a is rational as a Nash equilibrium. Subgame perfection does not eliminate this equilibrium because the game fragment starting at Player 2's move is not a subgame—it breaks Player 3's information set.

The other Nash equilibrium to this game is {A;a;[pL, (1 − p)R]}, with $p < \frac{1}{4}$. Player 3 credibly threatens Players 1 and 2 by playing R to force them to play A and a in this equilibrium. These strategies form a Nash equilibrium. If Player 2 plays a and Player 3 plays L with probability $p < \frac{1}{4}$, Player 1 prefers A to D. A secures him a payoff of 1, while D leads to a payoff of 3p for him. If Player 1 plays A and Player 3 plays R with probability $p < \frac{1}{4}$, a is Player 2's best reply. Her move is now on the equilibrium path. Choosing a gives her a payoff of 1, while d gives her a payoff of 4p, which is less than 1. Player 3's move is off the equilibrium path; when Players 1 and 2 play A and a, Player 3 never makes a move. Any move by Player 3 can be part of a Nash equilibrium.

Players 1 and 2 can collaborate to take advantage of Player 3's lack of knowledge of their moves in this game. If they can convince Player 3 to play L in the hope of obtaining the (3,2,2) outcome, they can exploit Player 3's inability to verify their moves and play A and d, leading to the (4,4,0) outcome. Consequently, Player 3 must play it safe by playing R, which keeps Players 1 and 2 honest. All three players are hurt by Player 3's lack of information. If we break Player 3's information set so that Player 3 can verify the prior moves of Players 1 and 2, this game has a Pareto-superior solution to the second Nash equilibrium.

Exercise 6.4: Verify that (D;a;R,L) is a subgame-perfect equilibrium for the game in Figure 6.1 if Player 3's information set is broken into separate nodes. (Note: The strategy above gives Player 3's moves for both nodes—R at the upper node and L at the lower node.)

The (D;a;L) equilibrium solves the information problem by allowing Player 2 to make a noncredible commitment to play a to Player 3. But Players 1 and 2 have an incentive to undermine that commitment, and Player 3 cannot verify that they have honored or broken that commitment. Why should Player 3 believe Player 2's commitment? To capture the intuition here, we introduce the concept of beliefs. We can then calculate how Player 3's beliefs about the moves of Players 1 and 2 drive her own move.

Definition: A set of **beliefs**, μ, for a game is a set of probability distributions with one distribution for each information set in the game.

Beliefs allow us to calculate expected utilities for each possible choice in a game. A belief for a given node is the conditional probability that the node is reached if the information set containing the node is reached during play of the game. Subgame perfection allows us to examine the rationality of moves within proper subgames but is powerless in the face of an information set that cannot be divided into a proper subgame. A set of beliefs specifies for each information set the probability that the player is at a given node in the information set for every node in the information set. Beliefs at a singleton information set must equal 1 by the laws of probability. For an information set with multiple nodes, the sum of the probabilities of all nodes in that information set must be 1. A player's expected utility is calculated by weighing its expected utility for each action at every node in the information set by the actor's belief that it is at that node. Actors then maximize expected utility at every information set, using their beliefs.

Example: Return to the game in Figure 6.1 and examine the rationality of the Nash equilibrium (D;a;L). Assume that Player 3's

belief that she is at her upper node is $\frac{2}{3}$ and her belief that she is at her lower node is $\frac{1}{3}$. The beliefs of Players 1 and 2 are both 1 because they have only singleton information sets. For Player 3, calculate expected utilities for each move given the above beliefs:

$$u(\text{Play } L) = (\tfrac{2}{3})(0) + (\tfrac{1}{3})(2) = \tfrac{2}{3};$$

$$u(\text{Play } R) = (\tfrac{2}{3})(1) + (\tfrac{1}{3})(0) = \tfrac{2}{3}.$$

Player 3 is indifferent between L and R and any mixed strategy of the two given these beliefs. Here, we choose the pure strategy L. Player 3 prefers L to R whenever her belief that she is at the lower node if her information set is reached is greater than $\frac{1}{3}$. There is a wide range of beliefs for which Player 3 prefers L to R.

One of the advantages of beliefs is that we can now check the rationality of any move in a candidate equilibrium, including those that are not contained in a proper subgame. Player 2's move of a was problematic before; now we can check whether that move is rational. The technique is similar to backwards induction. We trace the likely consequences of each of Player 2's available moves, and then calculate the expected utility of each. If Player 2 chooses a, she receives a payoff of 1. If she chooses d, Player 3 will choose L, and Player 2 receives a payoff of 4. Clearly, Player 2 prefers d to a. Thus a is not a rational move once beliefs allow us to carry out backwards induction through information sets.

For Player 1, the utilities of playing A and D are 1 and 3, respectively. (Recall that Player 2 plays a in the candidate equilibrium.) Player 1's move is rational given the other players' moves in the candidate equilibrium. Player 1's beliefs, like Player 2's, are irrelevant in calculating his expected utility because his information set is a singleton. However, Player 2 prefers d to a once beliefs allow us to perform a backwards induction from all information sets. Adding beliefs for Player 3 did not change the rationality of her move. Instead, this addition allowed us to see that Player 2's move was not rational. Beliefs permit us to evaluate all moves using expected utility calculations. I now define "rationality" with beliefs.

Definition: A pair of beliefs and strategies is **sequentially rational** iff from each information set, the moving player's strategy maximizes its expected utility for the remainder of the game given its beliefs and all players' strategies.

Exercise 6.5: Verify that (A;a;R) is sequentially rational for the game in Figure 6.1 for any set of beliefs where Player 3 places at least probability $\frac{2}{3}$ that she is at her upper node if her information set is reached.

We can describe the idea behind beliefs intuitively. A player who is uncertain about prior moves (i.e., at an information set with multiple nodes) creates

hypotheses about those prior moves. I say hypotheses here because the beliefs for one information set may involve speculation about prior, unobserved moves by several players, including moves by Chance. Player 3's beliefs at such a point depend upon conjectures about what both of the other players have done. These hypotheses could assert that one particular node has been reached in the information set. The beliefs then must place probability 1 on a node in the information set. They might assume that one of several nodes has been reached. Beliefs summarize what the player thinks has happened in the game before the current information set.

What beliefs are reasonable in the context of a given equilibrium? The beliefs should be based on the chance moves in the game and the other players' moves in the equilibrium whenever possible. Bayes's Theorem provides the mechanism for updating probabilities, and beliefs are just sets of conditional probabilities across the nodes of different information sets. The hypotheses a player uses to determine its beliefs should be based on the expectation of equilibrium behavior by the other players. As in Nash equilibrium, we assume that the players share a common conjecture that they are playing their equilibrium strategies. The players (and we) can calculate the probability that each node is reached from those equilibrium strategies. At a minimum, the beliefs must equal these conditional probabilities along the equilibrium path. Otherwise, the players' beliefs would diverge from their expectations about one another's behavior.

Example: Return yet again to the game in Figure 6.1. What beliefs does the (D;a;L) equilibrium produce for Player 3? We calculate the chance that her upper node is reached given that her information set is reached in this equilibrium. Player 3's upper node is reached if Player 1 plays A and then Player 2 plays d; her lower node is reached if Player 1 plays D. Denote "Player 3's upper node reached" by 3's un and "Player 3's information set reached" by 3's inf. We use Bayes's Theorem to calculate the probability that Player 3's upper node is reached if her information set is reached as follows:

$$p(3\text{'s un}|3\text{'s inf}) = \frac{p(A, d)p(3\text{'s inf}|A, d)}{p(A, d)p(3\text{'s inf}|A, d) + p(D)p(3\text{'s inf}|D)}$$

$$= \frac{(0)(1)}{(0)(1) + (1)(1)} = 0$$

Player 3 should not believe that she is at her upper node if her information set is reached; she must believe that she is at the lower node. When Player 1 plays D and Player 2 is committed to playing a, the only way Player 3's information set can be reached is her lower node.

Along the equilibrium path, we can calculate beliefs. But we cannot make such a calculation when a player must make a decision at an information set that has probability zero in an equilibrium. Instead, we allow the players to create a plausible hypothesis to explain what has happened. Something that should not happen in equilibrium has happened, and the players need some hypothesis to explain the defection. Using this hypothesis, each player can maximize its expected utility and continue playing. For now, we place minimal restrictions on such hypotheses.

> **Definition**: A **perfect Bayesian equilibrium** is a belief-strategy pairing such that the strategies are sequentially rational given the beliefs and the beliefs are calculated from the equilibrium strategies by means of Bayes's Theorem whenever possible.

I am being vague deliberately about beliefs off the equilibrium path when I say "whenever possible." Rather than stating technical definitions of what restrictions are placed on beliefs off the equilibrium path in perfect Bayesian equilibria, I discuss some of the issues here. First, the players continue to use the equilibrium strategies to update their beliefs after moves off the equilibrium path. Defection does not lead the players to abandon the common conjecture of equilibrium behavior. Instead, they assume that one defection does not increase the chance that other players will play "irrationally" off the equilibrium path. Second, in games with three or more players, we assume that if one player defects from its equilibrium strategy, the other players use the same conjecture about its defection. If they have the same beliefs prior to the defection, they must have identical beliefs after the defection. Third, players "cannot signal what they do not know." A defection by Player 1 does not lead Player 2 to change her beliefs about what Player 3 has done before 1's defection.

Perfect Bayesian equilibria, like Nash and subgame-perfect equilibria, always exist in mixed strategies.

> **Theorem**: Every finite n-person game has at least one perfect Bayesian equilibrium in mixed strategies.

This theorem is true because finite games always have perfect equilibria, and any perfect equilibrium is also perfect Bayesian.

There is no easy method for finding perfect Bayesian equilibria. I find the best technique is to think about how the game should be played, formulate a possible equilibrium, and check to see if the strategies are optimal given the beliefs and the beliefs follow from the strategies along the equilibrium path. Backwards induction can be very helpful in seeing what strategies might be in equilibrium and what beliefs are needed to sustain them. Alternatively, look for Nash equilibria, determine what beliefs follow along the equilibrium path, and see if the strategies are sequentially rational given the beliefs.

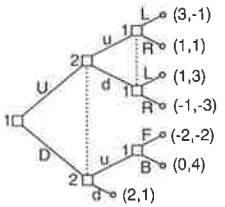

Figure 6.2 An Example of Perfect Bayesian Equilibrium

Example: Find the perfect Bayesian equilibria of the game in Figure 6.2. Specify a belief-strategy pair for this game as follows: (Player 1's move in the first node, Player 1's move in the upper information set, Player 1's move in the lower branch; Player 2's move: Player 1's belief that he is at the upper node of his information set in the upper branch if that information set is reached; Player 2's belief that she is at the upper node of her information set). We do not need to specify the beliefs for the singleton information sets. In general, I state strategies in a perfect Bayesian equilibrium before the beliefs. Strategies are specified using the same notation as for a Nash equilibrium; Player 1's complete strategy followed by the other players' strategies in order. The players' beliefs are listed after a colon; the beliefs of each player are separated from those of the other player(s) by semicolons. Individual moves within a player's strategy and beliefs for each information set within a player's set of beliefs are separated by commas.

Begin the analysis in the lower branch. Here, we have a classic example of subgame perfection. Player 1 would like to use the threat of F to force Player 2 to play d, but the threat is not credible. In equilibrium, Player 1 must play B at his lower node, and so Player 2 wants to play u if she is at her lower node.

In the upper branch, L is a dominant strategy for Player 1. Regardless of his beliefs at that information set, he prefers to play L. To see this, let p be Player 1's belief that he is at the upper node of the information set. Calculate expected utilities for both possible moves:

$$u(\text{Play L}) = p(3) + (1 - p)(1) = 1 + 2p;$$
$$u(\text{Play R}) = p(1) + (1 - p)(-1) = -1 + 2p$$

Playing L is better for any possible value of p. Anticipating that Player 1 will play L, Player 2 prefers playing d if she is at the upper node of her information set.

Now the trick in the game comes in—Player 2 does not know which node she is at when she must make her decision. She prefers playing d at the upper node and u at the lower node. We look for beliefs that make Player 2 indifferent between playing u and d, allowing her to employ a mixed strategy. Let q be Player 2's belief she is at the upper node. Then we have the following when she is indifferent between playing u and d:

$$\begin{aligned} \text{u(Play u)} &= \text{u(Play d)};\\ q(-1) + (1-q)(4) &= q(3) + (1-q)(1)\\ q &= \tfrac{3}{7}. \end{aligned}$$

Note that Player 2 anticipates Player 1's future moves when she calculates her utility for each move. The consequences of playing u is the (3,−1) outcome from the upper node because Player 1 will play L and the (0,4) outcome from the lower node because he will play B. If she chooses d, the outcomes will be (1,3) from her upper node and (2,1) from her lower node. If Player 2 believes she is at the upper node with probability $\frac{3}{7}$, then she is indifferent between playing u and d and can play any mixed strategy.

To produce these beliefs, Player 1 must play U with probability $\frac{3}{7}$ and D with probability $\frac{4}{7}$. Otherwise, Player 2's beliefs are not consistent with his equilibrium strategy. Because this information set must lie on the equilibrium path (both of Player 1's initial moves lead to Player 2's information set), Player 2's beliefs must be the same as the conditional probabilities each node is reached in equilibrium.

For Player 1 to mix his strategy in his first move, he must be indifferent between playing U and D. Player 2 can create this indifference by choosing a mixed strategy in her move. Let r be the probability that Player 2 chooses u in her move. Calculate Player 1's expected utilities for U and D and equate them:

$$\begin{aligned} \text{u(Play U)} &= \text{u(Play D)};\\ r(3) + (1-r)(1) &= r(0) + (1-r)(2)\\ r &= \tfrac{1}{4}. \end{aligned}$$

Once again, Player 1 anticipates Player 2's and his own future moves when calculating his utility for each strategy.

Putting all this together, $[(\frac{3}{7}\text{U},\frac{4}{7}\text{D}),\text{L},\text{B};(\frac{1}{4}\text{u},\frac{3}{4}\text{d}):\frac{1}{4};\frac{3}{7}]$ constitutes a perfect Bayesian equilibrium for this game. The beliefs follow directly from the strategies. Player 1 plays U with probability $\frac{3}{7}$ and D with probability $\frac{4}{7}$. Then Player 2's beliefs must be $\frac{3}{7}$ on the upper node and $\frac{4}{7}$ on the lower node. Similarly, Player 1's beliefs for his information set also follow directly from Player 2's strategy. In this game, Player 1 mixes his strategy to produce the beliefs that allow Player 2 to mix her strategy in a fashion that makes Player 1 indifferent at his first move, allowing him to mix his strategy. This interdependence of mixed strategies is common in these games. If either player deviates from the equilibrium strategy, the other player will take advantage of that defection.

Finally, we must check that no pure strategy equilibrium exists where Player 2 knows that she is at one of the two nodes in her information set. Player 1's moves later in the tree are fixed at L and B by the same logic as before. If Player 1 plays U for certain, Player 2 will believe she is at her upper node (consistency of beliefs again) and will play d. But then Player 1 would prefer to shift from U to D, so (U,L,B;d:1;0) is not a perfect Bayesian equilibrium. Similarly, (D,L,B;u:0;1) is not a perfect Bayesian equilibrium because Player 1 would like to change from D to U. If he does, Player 2 wants to change to d.

Beliefs allow us to judge the sequential rationality of moves from information sets with multiple nodes. In this example, Player 2's optimal move from her information set depends on her beliefs. In the example in Figure 6.1, beliefs allowed us to judge the rationality of Player 2's move at a singleton information set before Player 3's information set with multiple nodes. Sequential rationality judges the rationality of all moves in a game.

Exercise 6.6: For each of the Nash equilibria in Exercise 5.2 (page 130), determine which are perfect Bayesian equilibria. Find the beliefs that support each perfect Bayesian equilibrium.

Exercise 6.7: Find the perfect Bayesian equilibria for each of the games in Figures 6.3 through 6.5. Be certain to specify the beliefs and the strategies off the equilibrium path as well as the equilibrium behavior.

a) Find the Nash equilibria of the game in Figure 6.3 and compare them to the perfect Bayesian equilibria.

b) C denotes a chance move in the game in Figure 6.4. Find the perfect Bayesian equilibria.

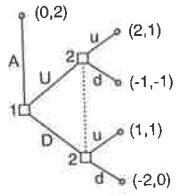

Figure 6.3 Exercise 6.7a

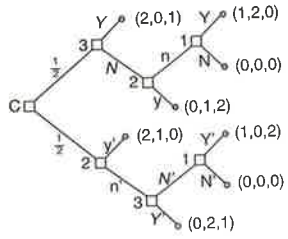

Figure 6.4 Exercise 6.7b

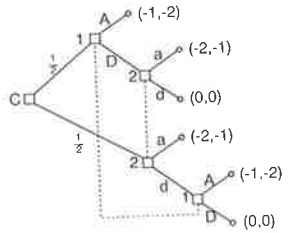

Figure 6.5 Exercise 6.7c

c) Each player has only one information set in the game in Figure 6.5. Each must choose its move without knowing the chance move that determines which player moves first. Find the perfect Bayesian equilibria.

Nuclear Deterrence

I now turn to some simple analysis of the strategic questions in nuclear war. All responsible parties agree that nuclear war would be an unparalleled disaster, but under what conditions might a government think about the unthinkable? To set the scene, I present the following greatly simplified discussion of some issues in nuclear strategy drawn from Powell 1990.

Some rational leaders might consider launching a nuclear first strike if that strike would disarm the other side, preventing any response (assuming that long-run ecological damage would not impose serious costs on the striking side). But during the Cold War, both the United States and the former Soviet Union had nuclear arsenals that made a first strike that disarmed the other side

highly improbable. From the mid-1960s on, each side had a secure second-strike capability; that is, both the United States and the Soviet Union could have responded to any initial nuclear strike with a devastating retaliatory strike, primarily from submarine-based missiles, but also from surviving land-based missiles. First strikes were deterred by this credible threat of retaliation. This case illustrates the general one: neither side will be willing to launch a first strike when such an attack will only lead to its own destruction through nuclear retaliation.

This conclusion has a disturbing side effect. It eliminates the use of nuclear weapons for extended deterrence—the protection of allies from external threats through nuclear threats. For example, during the Cold War, the United States threatened to use strategic nuclear weapons if the Soviet Union invaded Western Europe—but if such a nuclear first strike would necessarily have led to the devastation of the United States by Soviet nuclear retaliation, the threat of initiating nuclear war to defend Western Europe would not have been credible. For nuclear weapons to have political utility beyond the deterrence of nuclear war, both sides must believe there is some chance that nuclear war could start. Otherwise, the threat is hollow.

Schelling (1960) proposed one solution to this problem, the reciprocal fear of surprise attack.[1] Assume there is some advantage in striking first if nuclear war occurs: the side that strikes first is somewhat less devastated than the other. Both sides can still launch devastating second strikes. But it is better to strike first than second because the first strike takes out some of the other side's missiles. Each side might contemplate a first strike, not because it expected to win by attacking, but rather because it feared that the other side was preparing to attack and it wished to gain the first strike advantage for itself. These fears could build upon one another in a vicious circle, creating the reciprocal fear of surprise attack. Nuclear war might then be launched, not because either side thought it could win, but because each feared the other was about to launch an attack.

This argument places several restrictions on possible models. Neither side must know that the other side has committed itself to not attacking when it must decide whether to launch an attack itself. If neither side decides to attack, the status quo, the best outcome for both sides, should prevail. If a first strike is launched, the other side retaliates, but the side that strikes first suffers less. The game in Figure 6.6 is one model of the argument. The A and a actions are nuclear first-strike attacks, and the D and d actions delay the launching of a first strike. The a payoffs are for launching a first strike, and the r payoffs are for receiving such a strike and then retaliating. The difference between the two measures the first-strike advantage. The larger $(r - a)$ is, the greater the advantage to striking first. If neither player attacks, the status quo holds—the 0 payoff. We assume that striking first is preferable to receiving a first strike, but that no nuclear war is preferable to any nuclear war (i.e., $0 > -a_i > -r_i$

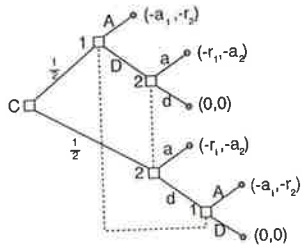

Figure 6.6 A Game with the Reciprocal Fear of Surprise Attack

and $0 > -a_2 > -r_2$). The chance move and information sets capture the idea that neither player knows whether the other is preparing a first strike when it must choose whether to launch a first strike of its own. Neither player knows whether delaying a strike ends the game at the status quo or gives the other player the opportunity to launch its own strike.

Specify an equilibrium of the game in Figure 6.6 by (Player 1's move; Player 2's move: Player 1's belief that he has the first move if his information set is reached; Player 2's belief that she has the first move if her information set is reached). The above game has three perfect Bayesian equilibria:

$$(A;a:1;1), (D;d:\tfrac{1}{2};\tfrac{1}{2}),$$

and

$$\left[\left(\frac{2a_2}{a_2+r_2}A, \frac{r_2-a_2}{a_2+r_2}D\right); \left(\frac{2a_1}{a_1+r_1}a, \frac{r_1-a_1}{a_1+r_1}d\right); \frac{a_1+r_1}{2r_1}; \frac{a_2+r_2}{2r_2}\right].$$

In the first equilibrium, each side attacks if it wins the draw because each knows that if it does not attack, the other side will attack in its turn. This equilibrium gives the reciprocal fear of surprise attack run amok. Each player attacks out of the fear that the other will attack if it does not. In the second equilibrium, neither side attacks because each knows that the other side will not attack in its turn. Here, we have "mutual confidence in restraint"; neither player launches an attack because they both believe the other player will not launch one.

To see that the first strategy-belief pair forms a perfect Bayesian equilibrium, consider a player's best reply given its beliefs and the other player's strategy. Call the player i for convenience. It believes that it has the initiative to strike first if it gets to move. Its utility for attacking is $-a_i$. If it delays its attack, the other player will attack, giving Player i a payoff of $-r_i$. Because $-a_i > -r_i$, it prefers attacking. Its beliefs follow from the players' strategies and Bayes's Theorem. Let "i mf" (or "j mf") stand for "Player i moves first" (or "Player j moves first"), which has probability $\frac{1}{2}$ based on the initial chance move. Let "i isr" stand for Player i's information set reached. If Player i has the first move, its information set is always reached, p(i isr|i mf) = 1. If Player j has the first move, i's information set is never reached in this equilibrium because j always attacks, p(i isr|j mf) = 0. Calculate the probability that Player i has the first move if its information set is reached:

$$p(\text{i mf}|\text{i isr}) = \frac{p(\text{i mf})p(\text{i isr}|\text{i mf})}{p(\text{i mf})p(\text{i isr}|\text{i mf}) + p(\text{j mf})p(\text{i isr}|\text{j mf})}$$

$$= \frac{(\frac{1}{2})(1)}{(\frac{1}{2})(1) + (\frac{1}{2})(0)} = 1.$$

In the third equilibrium, both sides play mixed strategies, with each side's probability of attacking increasing as the other side's first-strike advantage (r − a) decreases. If the third equilibrium seems bizarre, remember that each side's probability of attacking is chosen to make the other side indifferent between attacking and not attacking. One might think that the greater the first-strike advantage, the more attractive a first strike. However, there are two motivations for attacking in this model: to gain the first-strike advantage and fear of the other player's attacking in turn. The mixed strategy equilibrium offsets these two motivations. When the advantage from striking first is large, the motivation to strike first from fear must be reduced. Otherwise, the other player will always launch a first strike. The best reply is to attack against mixed strategies that use a higher probability of attacking than the equilibrium strategy does. When the opponent has a strong motivation to seize the first-strike advantage, you must try not to provoke it. Lowering the probability of launching one's own first strike lowers the level of provocation.

Exercise 6.8: Demonstrate that

$$\left(D;d:\frac{1}{2};\frac{1}{2}\right)$$

and

$$\left[\left(\frac{2a_2}{a_2+r_2}A, \frac{r_2-a_2}{a_2+r_2}D\right); \left(\frac{2a_1}{a_1+r_1}a, \frac{r_1-a_1}{a_1+r_1}d\right); \frac{a_1+r_1}{2r_1}; \frac{a_2+r_2}{2r_2}\right]$$

are perfect Bayesian equilibria of the game in Figure 6.6.

The model in Figure 6.6 formalizes the logic of the reciprocal fear of surprise attack. Both sides are willing to attack if each fears that the other side is about to attack. If you break both sides' information sets and play the game under perfect information, the reciprocal fear of surprise attack disappears. Each side knows then whether it is moving first or second when it must decide whether to attack. When it is moving second, it knows that the other side has not attacked. When it is moving first, it knows that the other player will know that it has not launched an attack when the other player moves. Only uncertainty about the other side's actions can create the reciprocal fear of surprise attack. If nuclear war were like tennis, where everyone knows who serves and in what order they serve, it would be less of a problem. Unfortunately, nuclear war is not tennis.

Exercise 6.9: Show that (D,D;d,d) is the only subgame-perfect equilibrium of the game in Figure 6.6 played under perfect information (read the strategy as Player 1's move if he moves first, Player 1's move if he moves second; Player 2's move if she moves first, Player 2's move if she moves second).

The model above provides no reason why either side would contemplate using nuclear weapons in the first place. Typically, nuclear strategists assume that some crisis would precede any thought of using nuclear weapons. A nuclear threat could be considered as a way to extort a favorable resolution of the crisis. In the model in Figure 6.6, there is nothing at stake between the two sides except nuclear war. If we add some stakes beyond the prevention of nuclear war to the model, each side has another option—to end the crisis by surrendering the stakes to the other side. I call this option Quit (abbreviated Q and q). The outcome of quitting the crisis is that the side that quits surrenders the stakes to the other. Winning the stakes is preferable to the status quo; surrendering the stakes is worse than the status quo but better than any nuclear war. Let the value of the stakes be s_i for Player i. Then $0 > -s_1 > -r_1 > -a_1$, and $0 > -s_2 > -r_2 > -a_2$. Figure 6.7 presents the game with this added option.

This game has only one perfect Bayesian equilibrium, (D;d:$\frac{1}{2}$;$\frac{1}{2}$) (using the same notation as for the previous game). Once we add the option of ending the crisis by surrendering the stakes, neither player has an incentive to attack because quitting the crisis is always preferable to starting a nuclear war. Consequently, the reciprocal fear of surprise attack disappears for both sides. If one side begins to fear that the other is planning to attack, it should quit the crisis instead of launching its own first strike. The logic of mutual assured destruction says that nuclear war, even when you strike first, is worse than any non-nuclear war outcome, including surrendering the stakes at hand. Thus the reciprocal fear of surprise attack should not occur. Not only should I surrender if I fear you are planning to attack, but I should also expect you to surrender if you fear I am planning to attack. Further, nuclear threats cannot be used in

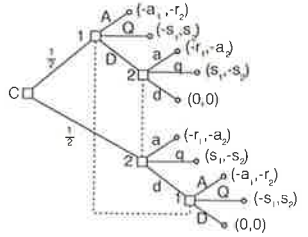

Figure 6.7 The Opportunity to Quit Added to a Game with the Reciprocal Fear of Surprise Attack

this model to coerce the other side to surrender the stakes. Both sides expect that the other side will not attack. There is no reason to surrender the stakes to eliminate the threat of war—that threat does not exist. Again we return to the argument that nuclear weapons cannot be used to defend other interests or extort concessions from the other side. Some real threat of nuclear war is necessary for either extended deterrence or nuclear extortion to be possible.

Exercise 6.10: Show that (D;d:$\frac{1}{2}$;$\frac{1}{2}$) is the only perfect Bayesian equilibrium of the above game.

a) First show that (D;d:$\frac{1}{2}$;$\frac{1}{2}$) is a perfect Bayesian equilibrium of the game.

b) Show that Q strictly dominates A (and q strictly dominates a). Consequently, A and a can never appear in a perfect Bayesian equilibrium strategy.

c) Show that once neither player ever attacks (play A or a), D strictly dominates Q (and d strictly dominates q). Thus only the given belief-strategy pairing can be a perfect Bayesian equilibrium.

The critical point of this example cannot be emphasized too strongly: *the results of a model depend upon the choices you give the players and how you structure them.* Breaking the information sets eliminated the problem, but that modification of the model seems implausible. The reciprocal fear of surprise attack depends upon each player's not knowing whether the other was preparing to attack. Eliminating that uncertainty denied a central premise of the argument. Adding the choice of quitting the crisis and sacrificing the stakes undermines the reciprocal fear of surprise attack, and that modification of the model does not violate the assumptions of the reciprocal fear argument. It may be that

the sides do not have the option to surrender in a specific situation. This variation does not demonstrate that the reciprocal fear of surprise attack can never occur. It does demonstrate that in this model the reciprocal fear of surprise does not occur when the option of surrendering is available. The only way to judge what are reasonable models is to understand the situation, build the models, and solve for their consequences.

Review

This chapter has introduced the ideas of beliefs and perfect Bayesian equilibrium. Beliefs allows us to perform backwards induction through information sets with multiple nodes. This is sequential rationality. Each player's moves must maximize its expected utility given its beliefs and the other players' strategies. Beliefs allow us to judge sequential rationality from information sets with multiple nodes. We weigh the utility of a move from each node in an information set by the probability that the moving player believes it is at that node when it moves. Unlike Nash or subgame-perfect equilibrium, sequential rationality allows us to check best replies at all information sets in a game.

Beliefs must be consistent with the players' equilibrium strategies whenever possible. Players use the equilibrium strategies and Bayes's Theorem to calculate the probability that each node in an information set with multiple nodes is reached. Bayes's Theorem combines prior beliefs and new information optimally to update probabilities. For information sets off the equilibrium path, players are free to make any conjecture about why defection from equilibrium occurred. However, they must share that conjecture and continue to use Bayes's Theorem and the equilibrium strategies after a defection.

Further Reading

Most statistics textbooks contain sections on Bayes's Theorem. DeGroot 1970 is a textbook on Bayesian decision theory. The model of preference for biased information is loosely adapted from Calvert 1985.

The discussion in this chapter draws on Selten 1975 and Kreps and Wilson 1982. Both of these papers are highly mathematical, very difficult reading, and immensely rewarding. I have drawn freely from their carefully crafted examples and terse discussions of their solution concepts. The three-player game is a well-known example from Selten 1975. The textbooks in noncooperative game theory provide more accessible treatments of perfect Bayesian equilibrium.

The section on nuclear war draws heavily on the work of Robert Powell, in particular Chapter Five of Powell 1990. The other chapters of Powell 1990 deal with other issues in nuclear deterrence.

Comparative Politics

Formal work in comparative politics typically draws on models of U.S. politics. The strongest area of application is the politics of advanced industrial democracies. Most democracies are multiparty ones. Laver and Schofield 1990 is an excellent place to begin reading the formal literature on multiparty democracy. Although it does not present formal models, it draws heavily on models. Multiparty systems change both electoral competition and government formation. Austen-Smith and Banks (1988) address the question of how voters' decisions are affected by their considering the effects of their votes on the government that forms. Austen-Smith and Banks (1990) and Laver and Shepsle (1990) model government formation and portfolio allocation. Baron (1991a) modifies the model of bargaining in legislatures discussed in Chapter Five to investigate whether moderate parties are more likely than others to be included in coalition governments. Baron (1993) shows that a multiparty system leads the parties to adopt distinct policy positions in campaigns and in office. Greenberg and Shepsle (1987) analyze how the possible entry of new parties leads existing parties to adopt different positions.

Variations in electoral and legislative rules across countries has also been modeled. Palfrey (1989) explains Duverger's law—that single-member districts with winner-take-all elections give rise to only two competing parties in each district. Cox (1990) considers how different electoral laws change party positions in elections. Huber (1992) compares the legislative rules of France and the United States, using formal models of legislative structure.

There are models of other issues in comparative politics. Wallerstein (1989, 1990) studies questions of union organization and corporatism. Pool (1991) examines the strategic incentives that official languages create. Kuran (1991) looks informally at the problems involved in judging when a population is ready for a revolution. Bates and Lien (1985) show that political leaders can gain by granting policy concessions and rights to their subjects. Geddes's (1991) model of political reform was covered in Chapter Four. Putnam's (1988) two-level game model connects domestic politics and foreign policy. Tsebelis (1990) uses linked models to analyze the dual internal and external incentives leaders face.