Game Theory for Political Scientists

Electoral Politics

The theory of elections addresses two related questions: why and for whom do citizens vote, and how do candidates compete in elections? Coughlin (1990) surveys the formal literature on elections. Aldrich (1993) reviews what we know about turnout from rational choice models. Both are good places to begin reading. You should also read the essay on comparative politics in Chapter Six. I survey research on multiparty elections there.

Candidate competition is typically modeled in a spatial framework. Enelow and Hinich 1990 is a collection of recent papers on spatial theory. McKelvey and Ordeshook (1986) examine candidate competition and voting behavior under limited information. Candidates use polls, and voters use endorsements, to assess their relative positions. Kollman, Miller, and Page (1992) examine party competition with a bounded rationality model.

Ledyard's (1984) model includes the candidates' choice of positions in a spatial competition within the model. His model is more general than the simplified version in the text. It is also more pessimistic about voter turnout because the benefits to a voter for determining the winner are smaller as the candidates' positions converge. Palfrey and Rosenthal (1985) expand on Ledyard's model to show its limitations as an explanation of why people in large electorates vote. Lohmann's (1993) model of political activism is closely related to these models of electoral turnout. She models political action as a signaling process.

A different approach to voting in mass elections is retrospective voting, where voters decide whether to retain the incumbent on the basis of his or her performance in office. Ferejohn 1986 is the first game-theoretic model of retrospective voting and is presented in Chapter Nine. Austen-Smith and Banks 1989 is a more advanced model of retrospective voting. Alesina and Rosenthal (1989) turn this argument on its head to claim that elections affect economic performance, rather than economics' driving elections. Retrospective voting is a common element of many models that address the formation of policy by officeholders. Morrow (1991) presents one example of how retrospective voting drives policy on arms control.

Models have been developed of many of the other facets of elections. Austen-Smith (1992) examines lobbying of elected officials by constituents. Lupia (1992) models the initiative process with a signaling model. Myerson and Weber 1993 is the latest in a line of studies examining how voting rules affect electoral competition. Snyder (1990) analyzes campaign contributions and spending patterns as a form of electoral competition.

Chapter Eight
Games of Limited Information and Restrictions on Beliefs

In games of limited or incomplete information, the players lack information about the exact nature of the game. They could be ignorant of the structure of the game, the other players' utility functions, or the exact consequences of their moves. The deterrence game in Chapter Seven is a game of limited information. The challenger did not know the resolve of the defender.

The players generally have different information about these fundamental uncertainties. Up to now, we have assumed that all aspects of the game are common knowledge. Information is common knowledge if all players know that information and all players know that the other players know that information, and so on. A player's **private information** is information known only by that player. The defender's resolve, its preference between war and backing down, was its private information in the deterrence game. Each player's private information determines its **type**. Each player is one of several types; a player knows its own type, and the other players know only the set of types that it could be. In the deterrence game, the defender was one of two types, resolute or irresolute. It knew which type it was, but the challenger only knew that the defender was one of those two types.

In games of limited information, the players make judgments about the other players' types. Their beliefs in a perfect Bayesian equilibrium represent these judgments. A player's beliefs give the probability of the other players' types. In the deterrence game, the challenger had beliefs about the defender's resolve at both of its moves.

A player's private information can affect its moves. A resolute defender always resists a threat in the deterrence game, while an irresolute defender resists only sometimes. A player's moves can reveal its private information to the other players. A player updates its beliefs about another player's type from the latter's moves. A player compares the likelihood that each type of the other player would make the observed move in the equilibrium they are playing. If one type of opponent would make a particular move while another would not, that move reveals the opponent's type. In the deterrence game, the challenger revises upward its belief that the defender is resolute after the defender resists a threat. Resolute defenders are more likely to resist.

The players can communicate their private information through their moves, creating the possibility of strategic deception, bluffing, and signaling. A

perfect Bayesian equilibrium of a game of limited information describes both the play of the game in the strategies and the change in the players' beliefs produced by the play of the game. An equilibrium of a game of limited information must specify the strategies and beliefs of all possible types of players. The equilibrium of the deterrence game specified strategies for both resolute and irresolute defenders. If we only specified the strategy of one type of defender, how could the challenger update its beliefs after the defender's move? It would know how likely one type of defender was to resist, but not the other. Because the challenger does not know the defender's type, an equilibrium must solve for both resolute and irresolute defenders. The challenger must consider what both types of defenders will do because it does not know which type it faces. So must we to model the challenger's decision problem.

On the equilibrium path, the equilibrium strategies and Bayes's Theorem define beliefs. But perfect Bayesian equilibrium allows beliefs off the equilibrium path to be chosen freely. This freedom can have profound consequences in a game of limited information. Consider the equilibrium of the deterrence game where the challenger never makes a threat. The defender's response and the challenger's decision as to whether to press the threat are off the equilibrium path. Those moves do not happen in this equilibrium. The challenger's updated beliefs determine its decision as to whether to press the threat. Because this action is off the equilibrium path, we can choose the challenger's belief that the defender is resolute to be sufficiently high that the challenger never presses its threat. If the challenger never presses its threat, then both types of defenders will resist any threat. Anticipating these consequences, the challenger never makes a threat. A player's beliefs about another player's type can determine its moves in a game of limited information. Beliefs off the equilibrium path can thus produce bizarre behavior on the equilibrium path in such games. Restrictions on beliefs are an attempt to deal with this problem.

In Chapter Three, we assumed that the game is common knowledge. We model a game of incomplete information while retaining this assumption with a clever trick developed by Harsanyi (1967–1968). A game of limited information begins with a chance move that determines each player's type. Each player is informed of its own type but not the other players' types. The chance move also produces the players' initial beliefs about the other players' types. All players know the probabilities of each player's type in the chance move. These probabilities are the players' initial beliefs about other players' types. A game of incomplete information consists of a set of identical games. Each of these games is played by a combination of the players' types, and the complete set is defined by all possible combinations of the players' types. The chance move determines which of these games is actually played. The players' uncertainty about the other players' types are modeled by information sets connecting its parallel nodes across games where its type is the same but

the other players' types differ. In the deterrence game, there are two types of defenders and one type of challenger. There are two possible games, one between the challenger and a resolute defender and one between the challenger and an irresolute defender. The chance move determines which of the two games is played. The defender knows which game is played, but the challenger does not. Its uncertainty is represented by the two information sets that connect its moves in the two games. This trick allows us to model a game of incomplete information with a game of imperfect information (i.e., information sets with multiple nodes).

This chapter presents five different games of limited information. It introduces signaling games, a particular class of games of limited information, with a game of pretrial bargaining. The informational role of congressional committees is modeled in a signaling game in the second model. Parties bargaining over an agreement typically do not know one another's value for an agreement. The third model is a simple signaling game of bargaining under such uncertainty. The fourth model examines how actions in a crisis could influence the acting nation's reputation. Deterrent actions are signals of future intentions in this model. The fifth model illustrates the problem of beliefs off the equilibrium path. I also include a section on "cheap talk," costless signals of strategic intentions, in this chapter. Although it is not a game of limited information, I can then compare costly and costless signals.

Games of limited information can be used to represent many different uncertainties. Players are often uncertain about one another's payoffs, as in the deterrence game. Congress is uncertain about the results of policies in the model of the workings of congressional committees in this chapter. One can create games where the players "do not know" their own preferences (Morrow 1994a). In these games, the players are uncertain about their preferences over the outcomes. They may know how to achieve each outcome but are uncertain about which outcome is best for them. Each player may have some private information about the value of the outcomes. Such games represent situations where the consequences of policies are very unclear.

Games of incomplete information allow us to analyze strategic communication. We treat communication as messages that the players can send in a game. Sending a message is just a move in the game. Messages may lead the players to change their future moves by altering their beliefs about the game. We can study how communication and actions affect one another in these settings. Some uncertainty is necessary for communication to affect the play of a game. Otherwise, the players do not have beliefs that could be altered by communication. Communication could help the players understand the other players' types in a game of incomplete information. Alternatively, communication could help the players coordinate on one of the multiple equilibria of a game. Imperfect information, that is, uncertainty about the history of the game, is necessary for coordination to play a role. Otherwise, each player can verify what moves the

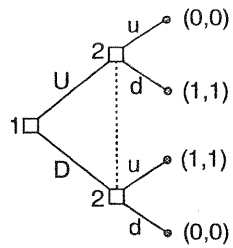

Figure 8.1 A
Coordination Game

other has made when it must move, and they do not need communication to signal their moves. The section on "cheap talk" analyzes such coordination.

Communication can have a dramatic effect on the equilibria of a game. It is difficult to say which equilibrium should occur if we do not specify if and how the players can communicate. Nash equilibria require a common conjecture to explain why the players hold mutual expectations about which equilibrium they will play. Communication is one way to provide such a common conjecture. Consider the game in Figure 8.1. If the players cannot communicate, we should expect that they will play the mixed strategy equilibrium, $[(\frac{1}{2}U, \frac{1}{2}D); (\frac{1}{2}u, \frac{1}{2}d)]$. If they can communicate, we expect that they will coordinate on one of the two pure strategy equilibria, (U;d) or (D;u). The outcome of this game depends on whether communication is allowed before the players choose their moves.

Formalizing communication in a game forces us to specify what communication is allowed. We can analyze how different forms of communication affect the equilibria and outcomes of the game. Communication between the players could also be accomplished through moves other than formal opportunities to exchange messages. Actions may indicate the intentions of the acting player, and so speak louder than words. Such communication is often called tacit communication because the players are inferring one another's meanings from their actions. If formal communication is allowed, the extensive form of a game should specify the opportunities to communicate and what messages can be sent. Tacit communication may occur in any dynamic game of limited information.

Signaling Games

Signaling games are a simple way to begin to think about communication. In a signaling game, Player 1, often called the sender, holds private information that affects the decision of Player 2, who is called the receiver. The first player has an opportunity to signal that information to the second player prior to the latter's decision. Consider a legal suit where the defendant knows whether it

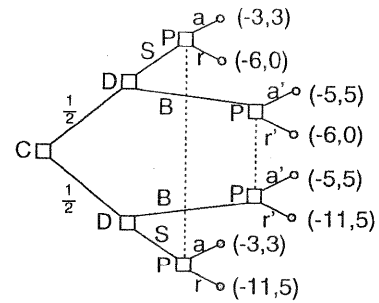

Figure 8.2 A Signaling
Game of a Lawsuit

has been negligent, and the plaintiff does not know. The defendant can offer either a small or a large out-of-court settlement to the plaintiff before trial. Going to trial is costly; the defendant prefers reaching a large settlement before the trial to bearing the costs of the trial. But a small settlement is better for the defendant than a large settlement. The settlement offered by the defendant could signal the strength of the defendant's case to the plaintiff, affecting the willingness of the plaintiff to accept the offer. If the settlement offer is rejected by the plaintiff, the case goes to trial, and the defendant's private information is revealed. If the defendant is negligent, the plaintiff is awarded an amount equal to the large settlement.

Figure 8.2 presents a game from Banks and Sobel 1987 that formalizes this situation. Player 1 is the defendant, here called Player D. Player 2 is the plaintiff, here called Player P. The defendant is not negligent in the top branch and negligent in the bottom branch. Moves B and S for the defendant are making a big or a small offer respectively. Moves a and r for the plaintiff represent accepting or rejecting the small offer; moves a′ and r′ represent accepting or rejecting the big offer.

(S,S;a,a′) is a Nash equilibrium of this game. Read the moves as (D's offer if not negligent, D's offer if negligent; P's response to a small offer, P's response to a big offer). The defendant prefers small offers if the plaintiff always accepts them. The plaintiff is always willing to accept small offers because the offer of 3 for certain is better than the gamble of rejecting it and going to court. P's expected utility for rejecting the offer is $\frac{1}{2}(0) + \frac{1}{2}(5) = 2\frac{1}{2}$, which is less than 3. Whether P accepts a big offer is irrelevant to this Nash equilibrium because D never makes a big offer in equilibrium.

> Exercise 8.1: Find the other two Nash equilibria of the above game. (Hint: There are only pure strategy Nash equilibria to this game. Why are there no mixed strategy equilibria on the equilibrium path?)

The next step is to find the beliefs that support each of the Nash equilibria. The combination of those beliefs and the strategies above form perfect Bayesian equilibria of the game. The plaintiff's beliefs specify how strongly the plaintiff believes the defendant is negligent after receiving an offer. Because the plaintiff has two information sets, two sets of beliefs must be specified—one for small offers and one for big offers. The beliefs are calculated from the strategies for any information set reached in the equilibrium. Off the equilibrium path, beliefs can be chosen freely. We will choose beliefs that support the plaintiff's strategy.

In the (S,S;a,a′) equilibrium, the information set after a small offer, S, is on the equilibrium path, and the information set for a big offer is off the equilibrium path. The plaintiff's beliefs after a small offer are calculated from the equilibrium strategies using Bayes's Theorem. Because both types of defendants, negligent and not negligent, always make a small offer, the plaintiff's beliefs must retain the initial probability of each type of defendant in the random move. Thus

$$
\begin{aligned}
&p(\text{D negligent}|\text{S}) \\
&= \frac{p(\text{D negligent})p(\text{S}|\text{D negligent})}{p(\text{D negligent})p(\text{S}|\text{D negligent}) + p(\sim\text{D negligent})p(\text{S}|\sim\text{D negligent})} \\
&= \frac{(\frac{1}{2})(1)}{(\frac{1}{2})(1) + (\frac{1}{2})(1)} \\
&= \frac{1}{2}.
\end{aligned}
$$

The defendant never makes a big offer in this equilibrium. What should the plaintiff believe about the defendant's negligence if she receives a big offer in this equilibrium? Because a big offer should not happen in this equilibrium, perfect Bayesian equilibrium allows the plaintiff to have any beliefs about the defendant's negligence after receiving a big offer. I calculate the range of beliefs that support the plaintiff's strategy of accepting all big offers. Any belief in this range will support this equilibrium. Let p be the plaintiff's belief that the defendant is negligent if he makes a big offer. Calculate the plaintiff's expected utilities for accepting and rejecting a big offer, set them equal, and solve for the belief that makes her indifferent between accepting and rejecting a big offer:

$$u_P(a'|B) = p(5) + (1 - p)(5) = 5,$$

$$u_P(r'|B) = p(5) + (1 - p)(0) = 5p,$$

and

$$a' I_p r' \text{ implies } p = 1.$$

If $p < 1$, the expected utility of accepting the big offer, 5, is greater than the expected utility of rejecting it, 5p, so she will accept a big offer. In words, the plaintiff must believe that the defendant is negligent for certain for her to reject a big offer. Even then, the plaintiff is indifferent between accepting and rejecting the offer. Any belief less than certainty, $p < 1$, means that the plaintiff accepts all big offers. Because we can choose beliefs off the equilibrium path freely, we can choose any $p \leq 1$, and the defendant will accept big offers.

The plaintiff's beliefs are $(\frac{1}{2},p)$ with $p \leq 1$ for the (S,S;a,a′) equilibrium, then. The beliefs are read (P's belief that D is negligent after receiving a small offer, P's belief that D is negligent after receiving a big offer).

> Exercise 8.2: Find beliefs that support each of the Nash equilibria you found in Exercise 8.1. Specify the beliefs by giving the conditional probability that D is negligent at each information set.

Both types of defendants make the same offer in all three equilibria. We call an equilibrium where all types play the same strategy a **pooling equilibrium**. A **separating equilibrium** occurs if the types play different strategies and the receiver can tell what type of sender it is facing after observing a move. A **semiseparating equilibrium** occurs when the types neither pool nor separate. The defendant's beliefs change after receiving the offer in a semiseparating equilibrium, but it cannot tell which type it is facing.

The game in Figure 8.2 has only pooling equilibria. The negligent defendants always want to try to convince the plaintiff they are not negligent so that the plaintiff will accept a small offer. The initial beliefs of the plaintiff allow negligent defendants to act as though they were not negligent. The plaintiff's initial belief that the defendant is not negligent is sufficiently high that the plaintiff will accept any offer that either type of defendant makes. The plaintiff learns nothing about the defendant's type in these pooling equilibria. Because all types make the same offer, the plaintiff's beliefs about the defendant's responsibility do not change after she receives the defendant's offer. With different initial beliefs, however, there are separating equilibria of this game.

> Exercise 8.3: Change the chance move in the signaling game in Figure 8.2 to a $\frac{1}{4}$ chance of the defendant's being not negligent and a $\frac{3}{4}$ chance of his being negligent.
>
> **a)** Show that (S,S;a,a′) and (S,S;a,r′) are no longer perfect Bayesian equilibria.
>
> **b)** Find a mixed strategy perfect Bayesian equilibrium of this game where negligent defendants make both types of offers and the plaintiff occasionally rejects small offers. (Hint: Remember

that mixed strategies are calculated to make the other player indifferent between playing its two strategies.)

The semiseparating equilibrium in Exercise 8.3 communicates information about the defendant's type to the plaintiff. Because some negligent defendants make big offers, the plaintiff's belief about the responsibility of the defendant changes after a small offer is made. It is less likely that the defendant is negligent. The plaintiff is less willing to go to trial as a result of the new information. But the plaintiff must still go to trial sometimes after receiving a small offer. If the plaintiff always accepts a small offer, then all defendants will make only small offers. The threat of a trial dissuades negligent defendants from making a small offer all of the time. At the same time, the plaintiff does not want to reject all small offers because nonnegligent defendants make small offers. The plaintiff is better off accepting a small offer from a nonnegligent defendant than rejecting it and losing at trial.

This example illustrates a general principle of equilibrium in signaling games. No type of sender can benefit in equilibrium by acting like a different type. Nonnegligent defendants make small offers, but negligent defendants make large offers sometimes. This principle is called **incentive compatibility**. In equilibrium, the actions of the different types must be compatible with their incentives. Otherwise, some type would prefer sending a different signal than that specified in the candidate equilibrium.

> **Definition**: A strategy pair $(\sigma;a)$ for signaling game S is **incentive compatible** iff for all player types t and t′,
>
> $$M_t\{\sigma(t); a[\sigma(t)]\} \geq M_t\{\sigma(t'); a[\sigma(t')]\},$$
>
> where M_t is the payoff function of type t, $\sigma(x)$ is the signal sent by type x according to strategy σ, and $a[\sigma(x)]$ is the action taken after observing message $\sigma(x)$ according to strategy a.

Pooling equilibria are always incentive compatible. All types send the same message in a pooling equilibria. The receiver takes the same action regardless of the sender's type; in other words, acting like a different type produces the same action. (Depending on the receiver's beliefs off the equilibrium path, there may be an advantage to sending a message that no type sends in the candidate equilibrium. This question is not one of incentive compatibility, though.) Incentive compatibility is important in testing separating equilibria. The receiver takes different actions after observing the sender's message in a separating equilibrium. Some types of senders benefit from imitation of other types when a strategy pairing is not incentive compatible. We check for incentive compatibility when looking for separating equilibria.

To conclude this section on signaling games, recall the model of political advice that produced a preference for biased information in Chapter Six, and compare it to a signaling game. The advisors in that model are not strategic actors. They produce recommendations without considering the effect of those recommendations on the decider's choice. If they were strategic actors, their recommendations would be chosen to influence the decider to make the choice that the advisor prefers. The next section covers a signaling model of the functioning of the committees of the U.S. Congress. The committee serves as an advisor to the whole body. It knows more about the consequences of policy than does the floor and reports bills that incorporate its superior information. If the committee prefers different policy outcomes from those preferred by the floor, the floor must be wary of accepting the committee's recommendations at face value.

The Informational Role of Congressional Committees

Signaling games provide us with a way to analyze communication between actors. We examine how receivers respond to the sender's signal. Earlier sections in Chapter Five presented models of congressional committees and rules. The argument from structure-induced equilibrium contended that the structure of Congress helped facilitate distributive deals over "pork barrel" programs. Committees provided their members with the ability to block legislation that opposed those members' interests and to send legislation to the floor that would pass unamended. The model of bargaining in legislatures indicated that closed rules were a way to expedite critical legislation when time pressure was high. This section presents a model of a third rationale behind the structure of Congress, drawn from Gilligan and Krehbiel 1987.

The effects of policies are uncertain. The common interest of all members in writing policy that serves their interests (leaving aside the question of how such interests are defined) calls for collecting information to reduce uncertainty. Such information collection requires members who are experienced in particular areas of policy. Committees exist to collect information, develop expertise in areas of policy, and write legislation for the floor based in part on their expertise and information. Legislative rules are used by the floor both to foster expertise and to control the influence of committees. Closed rules favor legislation written by committees, which encourages members to specialize in particular areas of policy. Open rules provide the floor with a way to amend legislation that reflects the particular interests of the committee at the expense of the floor.

The simplified model in this section allows the actors to "change their preferences." Outcomes combine both policy and factors beyond the actors' control. You can think of these exogenous factors as the details of the problem

that policy is trying to remedy. Without a detailed knowledge of the specific issues, policies may produce outcomes very different from their intended consequences. This combination of chosen policy and exogenous effects is a way to represent the uncertainty of policy in a formal model. Members have ideal outcomes that they wish to achieve. But without knowledge of the exogenous factors that influence outcomes, actors cannot determine what policy they prefer. Policy preferences can change as members learn about the exogenous factors that influence the outcomes that policies produce. In this sense, the actors' preferences can "change."

Because policy preferences depend upon knowledge of exogenous factors, members have an incentive to gain the expertise needed to judge the consequences of policies. But gaining expertise is costly, and no member can be an expert on all issues.[1] Committees exist to divide the burden of acquiring such expertise. But the preferences of committees may diverge from those of the floor. The floor needs not only to inspire committees to make the effort to collect information but also to control their particular interests. Legislative rules provide one way to do so.

The model has two actors, a committee, C, and the floor, F. Each actor represents the median position of its body. The median voter concept allows us to simplify the choices of each body to that of a single actor. The game analyzes one issue. The issue is represented by a single dimension. The status quo policy is y_{SQ}. Fix the floor's ideal point to be 0 and the committee's ideal point to be $x_C > 0$. A bill gives a policy, y. The outcome combines the final policy with exogenous effects. These exogenous effects are represented by ϵ, a random variable distributed uniformly on $[-1,1]$, that is, distributed so that all values between -1 and 1, inclusive, are equally likely. The outcome is $y + \epsilon$. The players' utility for the outcome is the opposite of the squared difference between the outcome and their ideal point. The further the outcome is from an actor's ideal point, the less preferred it is.

The game begins with a chance move to determine the value of ϵ, which is not revealed to either player. The committee then decides whether to make the effort to collect information on an issue. Collecting information costs the committee c_c and reveals the true value of ϵ to it. The committee then proposes bill y to the floor. Unlike the model using structure-induced equilibrium, this model specifies that the committee must report a bill to the floor. Under the open rule, the floor can then amend the bill freely; call the amended bill y'. The floor chooses between the bill and the status quo. Finally, the outcome is revealed, and the players receive their payoffs. The time line of this game is as follows:

1. Chance determines ϵ, with uniform distribution on $[-1,1]$.

2. C decides whether to observe ϵ. If so, it pays cost c_C from its final payoff. F is informed of whether C decided to observe ϵ.

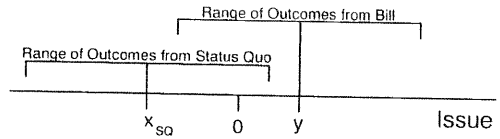

Figure 8.3 Comparing the Status Quo and Policy y with an Uninformed Committee

3. C proposes bill y. F updates its beliefs about ϵ after receiving y.

4. If the game is being played under the open rule, F can amend y to y'. If it is being played under the closed rule, this move is not available to F.

5. F chooses between y (or y') and y_{SQ}. Call its selection Y.

6. ϵ is revealed to both players. C receives $-(Y + \epsilon - x_C)^2 - c_C$ if it observed ϵ and $-(Y + \epsilon - x_C)^2$ if it did not. F receives $-(Y + \epsilon)^2$.

This model allows us to analyze how the incentives to use committees to collect information vary with the rules, the status quo, and the divergence between the floor and the committee. The greater x_C is, the greater the divergence between the floor's and the committee's ideal points relative to the uncertainty of the issue. ϵ is bounded between -1 and 1. If x_C is large, say, above 10, the uncertainty on the issue is small relative to the divergent preferences. If x_C is small, say $\frac{1}{4}$, the uncertainty is relatively large. Incentives to use committees for information should be greater in the second case than in the first.

I solve the model under the closed rule and leave the open rule to accompanying exercises. Find an equilibrium by breaking the problem into the cases where C chooses to observe ϵ and those where C chooses not to observe ϵ. We can compare the two cases to determine when C prefers to gain expertise on the issue.

Begin with the case where C does not collect information. The analysis of this case parallels the model of structure-induced equilibrium in Chapter Five. F must choose between bill y and x_{SQ}. Because neither F nor C is informed about ϵ, both still believe that ϵ is uniformly distributed on $[-1,1]$. The mean of ϵ is 0, and F prefers y over x_{SQ} if y is closer to 0, F's ideal point, than x_{SQ} is. Figure 8.3 illustrates this logic. All points in each range of outcomes are equally likely. F chooses the range whose center is closer to its ideal point. In Figure 8.3, F prefers the bill to the status quo.[2]

The committee reports the bill to the floor that is closest to its ideal point and can pass the floor. Any bill y closer to F's ideal point than x_{SQ} passes on the floor. Because the committee is uninformed about the true value of ϵ, it prefers any bill to the status quo that is closer to its ideal point. If $x_{SQ} \leq -x_C$, C proposes x_C as the bill. If $-x_C < x_{SQ} < 0$, C proposes $-x_{SQ}$. If $0 \leq x_{SQ} \leq x_C$, C proposes a bill that will not pass the floor, x_C, for example. If $x_C < x_{SQ}$, C

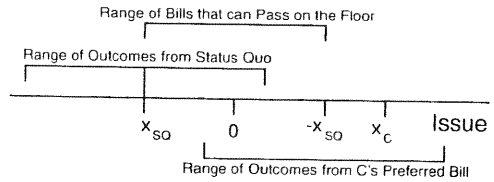

Figure 8.4 What Bills Will Pass on the Floor with an Uninformed Committee

proposes x_C. Figure 8.4 illustrates the second case here. Any bill between x_{SQ} and $-x_{SQ}$ passes the floor. The floor proposes the bill closest to its ideal point, x_C.

Exercise 8.4: Find the equilibrium under the open rule when the committee does not collect information.

a) Find the set of bills the floor prefers to the status quo.

b) To what bill y', if any, will F amend y?

c) What bill y does C report to F? Does it matter?

What are the consequences of the uninformed committee under the closed rule? The closed rule gives the committee significant proposal power. When the status quo is far from both the floor's and the committee's ideal points ($x_{SQ} \leq -x_C$ or $x_C \leq x_{SQ}$), the committee can achieve its ideal point as the approved bill. When the status quo is close to the floor's and the committee's ideal points ($-x_C < x_{SQ} < x_C$), the committee can block change or shape change to serve its interests at the expense of the floor. The closed rule then provides an advantage to the committee. However, the final outcome will not be its ideal point except in the rare case that $\epsilon = 0$. If $\epsilon < -x_C$ or $\epsilon > 0$ and the floor and the committee knew its true value, both would prefer some other bill to x_C. The final outcome, $x_C + \epsilon$, falls outside the set between their ideal points. Although passing x_C is optimal ex ante facto (when proposed and passed), it may be inefficient ex post facto (after the true value of ϵ is revealed). Both parties' lack of information about the consequences of policy limits their ability to choose the policy that leads to the outcomes they desire.

Exercise 8.5: Find the outcome under the open rule when the committee is uninformed. Who benefits relative to the closed rule? Find the values of ϵ where both the committee and the floor would prefer to substitute some other bill if they knew the true value of ϵ.

The uninformed case parallels the model of structure-induced equilibrium in Chapter Five. There are no signaling issues because the committee is uninformed. But the case where the committee is informed is more difficult. Signaling issues arise because the floor can learn from the bill the committee reports. Start by looking for a separating equilibrium. Such an equilibrium can be supported only if the floor prefers the bill to the status quo and the committee offers a different bill for every value of ϵ. The floor knows ϵ when it must vote in a separating equilibrium. The value of ϵ gives the committee's type. By the definition of a separating equilibrium, the floor can determine the committee's type from the bill the committee proposes. The floor then also knows ϵ when it votes.

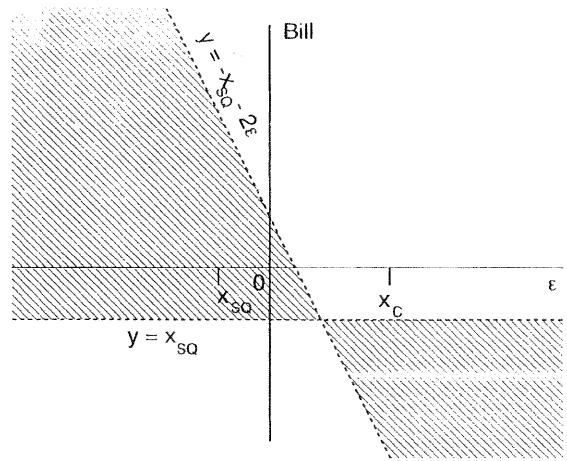

Figure 8.5 What Bills Will Pass on the Floor When an Informed Committee Signals Its Information

Figure 8.5 shows what bills the floor will approve. The horizontal axis gives the value of ϵ, and the vertical axis gives the proposed bill. The floor passes any bill in the shaded area between the two dashed lines in Figure 8.5. The equation of the horizontal line is $y = x_{SQ}$; the floor is indifferent between the status quo and a bill affirming the status quo. The equation of the diagonal line is $y = -x_{SQ} - 2\epsilon$; that bill produces a final outcome of $-x_{SQ} - \epsilon$, which is the opposite of the outcome the status quo produces, $x_{SQ} + \epsilon$.

Given this voting rule for the floor, what bill should the committee report? Figure 8.6 illustrates the logic. The solid diagonal line gives the bill that produces the committee's ideal point given ϵ. The equation of this line is $y = x_C - \epsilon$. The committee would like the final bill to be as close to this line as possible. When $\epsilon < -x_C - x_{SQ}$ or $\epsilon > x_C - x_{SQ}$, F will pass C's preferred bill, so C should propose that bill. When $-x_C - x_{SQ} < \epsilon < x_C - x_{SQ}$, the floor will not pass the committee's preferred bill. C then proposes the closest bill to its preferred bill that will pass the floor. The thick line in Figure 8.6 gives the bill that the committee should propose in this candidate separating equilibrium. F can determine the value of ϵ by working backwards from the reported bill.

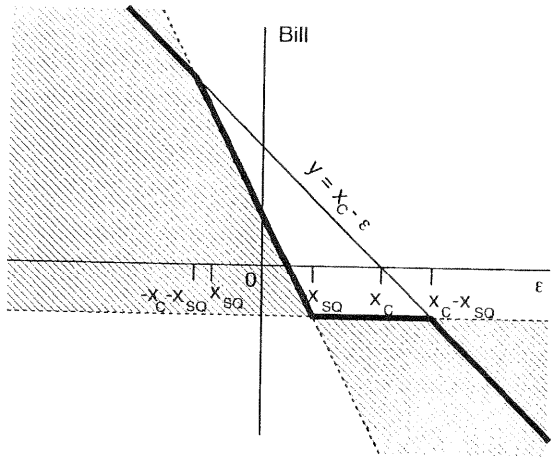

Figure 8.6 What Bills an Informed Committee Would Like to Report to the Floor

But this strategy is not incentive compatible when $-x_C - x_{SQ} \leq \epsilon \leq x_C - x_{SQ}$. The committee prefers sending a different bill than the bill given by this strategy. For example, let $\epsilon = -x_{SQ}$. The strategy proposes bill x_{SQ}, which produces a final outcome of $0 < x_C$. But C is better off proposing $x_C + x_{SQ}$. F infers that $\epsilon = -\frac{1}{2}x_C - x_{SQ}$ after observing that bill if the types separate according to the strategy for proposing bills above. F then accepts $x_C + x_{SQ}$, believing that that bill produces final outcome $\frac{1}{2}x_C$ and the status quo produces $-\frac{1}{2}x_C$. The end result of imitating the type where $\epsilon = -\frac{1}{2}x_C - x_{SQ}$ is x_C. That result is better for the committee than the outcome of 0 that results from the bill the separating strategy suggests. For all $-x_C - x_{SQ} \leq \epsilon \leq x_C - x_{SQ}$, similar bills can be found that produce x_C by tricking the floor about the value of ϵ.

The intuition here is simple. The bill that the committee reports is a credible signal of the exogenous effects on outcomes, ϵ, when those effects are large relative to the differences in preferred outcomes between the floor and committee. When the exogenous effects are large, the floor prefers accepting the committee's ideal outcome to the outcome the status quo policy produces. When the exogenous effects are small relative to their differences in preferred outcomes, the floor cannot trust the reported bill to reflect the exogenous effects accurately. If it does, then the committee can exploit that trust to gain better final outcomes for itself at the expense of the floor.

Exercise 8.6: Examine the possibilities for a separating equilibrium under the open rule with an informed committee. The open rule differs from the closed rule because the committee's bill can be amended by the floor to anything the latter chooses.

a) What amended bill will the floor pass for a given value of ϵ if the floor knows the true value of ϵ?

b) Consider the incentives of the committee if its proposed bill separates and allows the floor to know the true value of ϵ. Is separation ever incentive compatible for the committee? (You may assume that the committee's bill simply gives the value of ϵ directly.)

A separation of types cannot be supported for $-x_C - x_{SQ} \leq \epsilon \leq x_C - x_{SQ}$. Within this range, there must be a pooling equilibrium. Further, all types of committees in this range offer the same bill, and the floor always rejects that bill. Assume that there is a semiseparating equilibrium such that all types of committees between x_0 and x_1, with $-x_C - x_{SQ} \leq x_0 \leq \epsilon \leq x_1 \leq x_C - x_{SQ}$, report the same bill B, and F approves that bill. The committee of type x_0, the committee of type x_1, and the floor prefer bill B to x_{SQ} as the policy. If committees of types x_0 and greater prefer the bill, its outcome, $B + x_0$, must be less than x_C but greater than $x_{SQ} + x_0$. Otherwise, all types greater than x_0 prefer x_{SQ} to B. But then all types between $-x_C - x_{SQ}$ and x_0 also prefer the bill to the status quo. Then $x_0 = -x_C - x_{SQ}$. If a committee of type x_1 prefers the bill, then $x_C - x_{SQ} - x_0 \geq B + x_0 - x_C$ (the outcome of the bill is closer to the committee's ideal point than the outcome of the status quo policy). The two outcomes, $B + x_0$ and $x_{SQ} + x_0$, must bracket x_C. Otherwise, types greater than x_1 prefer B to x_{SQ}. This inequality implies that

$$B \leq 2x_C - 2x_1 - x_{SQ}.$$

Equality holds when $x_1 < x_C - x_{SQ}$. The floor's expected utility for the bill must be greater than its expected utility for the status quo. It prefers the bill when the midpoint of the possible outcomes of the bill is closer to its ideal point, 0, than the midpoint of the possible outcomes of the status quo.[3] This implies that

$$B + \frac{x_0 + x_1}{2} \leq -x_{SQ} - \frac{x_0 + x_1}{2}.$$

Simplifying this inequality, we obtain

$$B \leq -x_{SQ} - x_0 - x_1.$$

These two inequalities imply that $x_1 = x_C - x_{SQ}$. If $x_1 < x_C - x_{SQ}$, then $B = 2x_C - 2x_1 - x_{SQ}$. Substitute this expression for B in the second inequality to obtain $x_1 - x_0 \geq 2x_C$. But $x_0 = -x_C - x_{SQ}$, so $x_1 \geq x_C - x_{SQ}$. Then all types of committees between $-x_C - x_{SQ}$ and $x_C - x_{SQ}$ pool by reporting the same bill.

The floor rejects any bill presented by the range of committees that pool. If a committee reports such a bill, the floor believes that all values of ϵ between $-x_C - x_{SQ}$ and $x_C - x_{SQ}$ are equally likely. The range of outcomes the floor expects from the status quo after observing a pooling bill is $-x_C$ to x_C. No bill can beat that range of outcomes for the floor given the floor's uncertainty about the true value of ϵ.

I have shown so far that separation is incentive compatible for all types less than $-x_C - x_{SQ}$ or greater than $x_C - x_{SQ}$, and that all types within that range offer the same bill and have their bill rejected by the floor. These two observations create one final problem in finding a complete equilibrium. When types in the pooling range know that their bill will be rejected, they have an incentive to imitate a type in the range $-3x_C - x_{SQ}$ to $-x_C - x_{SQ}$. Consider $\epsilon = -x_{SQ}$ as an example. Under the pooling strategy, a committee of this type expects an outcome of 0. By reporting a bill of $2x_C + x_{SQ}$ instead, the floor will believe that $\epsilon = -x_C - x_{SQ}$ and approve that bill over the status quo (the final outcomes given the false belief are equally distant from 0). But the true final outcome under this deception is $2x_C$, which is as good for the committee as the outcome of the status quo policy.

The floor prevents this deception by refusing to draw the inference the separating equilibrium suggests for any bill between $4x_C + x_{SQ}$ and $2x_C + x_{SQ}$. If it does so, the committee can exploit that inference to its own advantage when ϵ is between $-x_C - x_{SQ}$ and $x_C - x_{SQ}$. Separation and complete communication of the committee's information is possible only when $\epsilon < -3x_C - x_{SQ}$ or $\epsilon > x_C - x_{SQ}$. The full equilibrium is as follows:

> Committee's Bill: If $\epsilon < -3x_C - x_{SQ}$, propose $x_C - \epsilon$. If $-3x_C - x_{SQ} \leq \epsilon < -x_C - x_{SQ}$, propose $4x_C + x_{SQ}$. If $-x_C - x_{SQ} \leq \epsilon < x_C - x_{SQ}$, propose B, with $4x_C + x_{SQ} < B < x_{SQ}$. If $\epsilon \geq x_C - x_{SQ}$, propose $x_C - \epsilon$.
>
> Floor's Beliefs: If $B > 4x_C + x_{SQ}$, $\epsilon = x_C - B$. If $B = 4x_C + x_{SQ}$, ϵ is uniformly distributed on $[-3x_C - x_{SQ}, -x_C - x_{SQ}]$. If $4x_C + x_{SQ} > B > x_{SQ}$, then ϵ is uniformly distributed on $[-x_C - x_{SQ}, x_C - x_{SQ}]$. If $B < x_{SQ}$, $\epsilon = x_C - B$.
>
> Floor's Vote: Accept B if $B \geq 4x_C + x_{SQ}$ or $B < x_{SQ}$. Reject B otherwise.

This equilibrium illustrates how committees can be used to collect information in the floor's interest. It also shows the limits of a committee system, when it works best, and whose interests it serves. The committee fully reveals its information when the results of the status quo policy are far from both the floor and the committee's ideal points. This range is asymmetric; accurate signaling is easier when the committee's private information works against its particular interest (i.e., when $\epsilon > x_C - x_{SQ}$). The committee has an incentive to overstate the effects of ϵ to shift the outcome of policy away from the floor's ideal point and towards its own ideal point. The floor deters such overstatements by ignoring all bills that might be used for deception. Of course, this response hurts both parties by preventing some bills where both parties would be better off with honest revelation (i.e., $-3x_C - x_{SQ} < \epsilon < -x_C - x_{SQ}$). The committee is rewarded for revealing its information when ϵ is large in either direction. It gets to set policy to produce its ideal point, and the closed rule prevents the floor from amending the bill in its own interest. But this protection benefits both parties. The committee would not reveal its information without the reward. The floor is better off with the committee's ideal point than with the outcome of the status quo policy.

The committee system is more effective at communicating information when the committee and floor have similar preferences over the outcomes. The closer to 0 that x_C is, the greater the range of values where the committee reveals its information. The smaller the cost of the reward to the floor, the smaller the incentive for the committee to try to deceive the floor, and the more likely the floor is to grant the reward. All of these incentives reinforce the committee's incentive to reveal information truthfully. This result parallels the preference for biased information discussed in Chapter Six. Signaling is most effective when the sender and receiver have the same preferences. When their preferences differ, the sender has an incentive to try to deceive the receiver. The receiver knows this and consequently discounts the sender's signals.

Informed committees reduce some, but not all, of the inefficiencies that uncertainty introduces. There are many cases where both the committee and the floor could be made better off after the fact when the committee was uninformed. The final outcome is inefficient when it falls outside of 0 and x_C. Inefficiency occurs with an informed committee when $-3x_C - x_{SQ} < \epsilon < -x_C - x_{SQ}$. The final outcome is $4x_C + x_{SQ} + \epsilon > x_C$. In all other cases, the final outcome falls between 0 and x_C, typically falling at x_C.

> Exercise 8.7: Exercise 8.6 showed that separation cannot be supported under the open rule. Find a semiseparating equilibrium under the open rule. Divide the range of possible ϵ into intervals between $x_0(=-1), x_1, \ldots, x_n(=1)$. C signals in which interval that ϵ lies. F believes then that C is uniformly distributed on that interval.
>
> **a)** Find the bill the floor approves if it believes that ϵ is uniformly distributed on $[x_i, x_{i+1}]$.
>
> **b)** Find the indifference condition for C when $\epsilon = x_i$. Set C's value for the bill that is approved if it signals that ϵ lies between x_{i-1} and x_i equal to its value for the bill approved if ϵ lies between x_i and x_{i+1}.

We can use the relationship among the x_i's found in part b) to solve for the exact limits of each interval.

Under the open rule, the communication of information from the committee to the floor is not as efficient. The floor always chooses a bill to suit its interests over the committee's. The committee cannot use its informational advantage to shift legislation in its interest, so its incentive to reveal its information is reduced. What information the committee reveals reflects only the broad outlines of the likely outcomes of policy. The information the committee reveals serves the interests of both the floor and the committee. The greater the difference in their interests (i.e., the bigger x_C), the less information transmitted.

This observation suggests another rationale for restrictive amendment rules on the floor. Such rules advance the motivation for a committee to seek information on the consequences of policy in uncertain areas. Such information benefits the floor by avoiding legislation with disastrous consequences. Closed rules provide a benefit to committee members to encourage them to collect and reveal information about the likely consequences of policy. Open rules increase the ability of the floor to discipline the bills that committees report, but at the cost of reducing the committee's incentive to reveal what it knows.

Contrasting this model with the structure-induced–equilibrium model of the committee system in Chapter Five provides a test of these competing hypotheses. In the latter model, rules were a way to enforce distributive bargains among members. Members received committee positions that fit their particular interests and promoted legislation that suited those interests. Closed rules helped protect that legislation from the general interest of the floor. In the informational model, committees also received particular benefits from the closed rule. But the two stories differ about the composition of committees they predict. The structure-induced–equilibrium model suggests that committee members should look different from the general floor. Members seek committee assignments that reflect their particular rather than general interests. The informational model says committee members should not differ from the floor membership. Signaling is most effective when there is little difference between the floor's and committee's preferred outcomes. The critical test here is to see whether the preferred outcomes of committee members regularly differ from those of floor members on the issues that a committee covers.

I am not going to go into the difficulties of conducting such a test here. They are significant; see Snyder 1992 for one problem. My point is that the models helped us understand the consequences of both models and this critical test between them. Both of these views are compelling when described in general terms. Both probably represent some aspect of the reality of congressional politics. But without the models, we would be hard pressed to see the consequences of the strategic logic of each view. There is much evidence used to support one view or the other that does not allow us to separate between the two views. In both models, for example, committees receive preferential treatment of their bills under the closed rule, and that preferential treatment leads to distributional gains for the committee at the expense of the floor. This is how models help us sort out the evidence that might allow us to judge which view matches reality.

Bargaining under Incomplete Information

Bargaining under complete information tends to be quite strange. If an agreement is possible, then the parties must reach it if either party faces a final choice of accepting the last offer or ending bargaining and the other side knows it. If the order of offers and responses is structured into a game with complete and perfect information where each side can make and accept any offer at any move, the sides should reach agreement immediately. The Rubinstein model in Chapter Five has both of these features. Both of these observations differ from how bargaining actually occurs. Most observers of bargaining find cases where the parties fail to reach an agreement, even when an agreement acceptable to both sides exists. Bargaining fails in these cases because one or both sides played too "hard" in the bargaining, leading the other side to stop bargaining in the belief that a bargain was impossible. Moreover, each side tends to resist making concessions in order to help preserve its position and achieve the best possible deal.

The failure of bargaining models under complete information to explain these bargaining behaviors is disturbing and leads us to explore bargaining under incomplete information. The insight underlying these models is simple: the essence of bargaining is communication, and when the players already know everything about the game, there is no need for them to communicate and therefore no bargaining. These models examine how players should bargain when they possess private information about the attractiveness of different bargains. Offers and rejections operate as signals to the other player, influencing its willingness to accept or reject offers.

There are many models of bargaining under incomplete information. Most of them are technically complex, and all of them work on very simple bargaining situations. This field has a long way to go to capture what most of us think of as bargaining. Nevertheless, it is one of the most promising areas in game theory, both in theory and application.

To illustrate these models, I present a simple model of bargaining under incomplete information. I begin with a complete information model. After solving it, I add some uncertainty. There are two players, a buyer (B) and a seller (S). The buyer offers one of two prices, high (7 units) or low (5 units), to the seller. The seller can then accept or reject the offer. If the initial offer is rejected, then the buyer can make a final offer of 7 or 5. The players have a discount

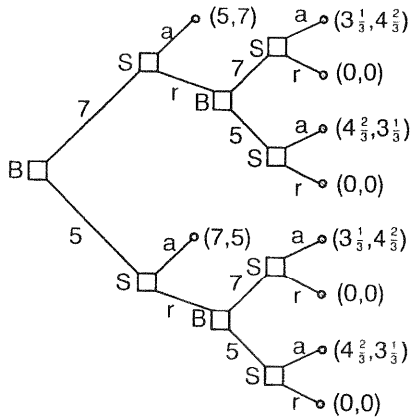

Figure 8.7 A Bargaining Game Where the Buyer Can Make Two Offers

factor of $\frac{2}{3}$ (if they make a trade in the second round, they receive only two-thirds of the value), so they prefer a trade in the first round to the same trade in the final round. The object they are bargaining over has 0 value to the seller and value 12 to the buyer at the end of the game. Figure 8.7 gives the extensive form of this game. The payoff for the buyer is given first, followed by the seller's payoff. The actions are labeled 7 and 5 for the offers and a and r for accept and reject for the seller's reaction to an offer. I have not differentiated all the offers and responses.

Exercise 8.8: Find the subgame perfect equilibria of the bargaining game in Figure 8.7 under complete information.

Now add some uncertainty to the bargaining. Let there be two possible types of sellers, one that places no value on the object, called soft sellers, and another that places value 6 on the object, called tough sellers. Tough sellers never accept low offers. The tough seller's payoff is its net gain from the game. If a tough seller replaces the soft seller shown in Figure 8.7, the payoffs change as follows: 7 goes to 1, 5 goes to -1, $4\frac{2}{3}$ in the second round (corresponding to an offer of 7) goes to $\frac{2}{3}$, $3\frac{1}{3}$ goes to $-\frac{1}{3}$, and 0 remains as 0. I state the game with incomplete information as the time line below rather than as an extensive form because it is easier to analyze in time line form:

1. Chance determines the seller's payoffs with probability $\frac{1}{4}$ that $r_S = 6$ (the seller is tough) and probability $\frac{3}{4}$ $r_S = 0$ (the seller is soft). The seller (S) is informed of its payoffs.

2. Buyer (B) offers either a high (7) or low (5) price. Call this offer Ω_1.

3. S accepts or rejects Ω_1. If it accepts, the game ends with $u_B = 12 - \Omega_1$ and $u_S = \Omega_1 - r_S$.

4. If S rejects Ω_1, B offers Ω_2 with either a high (7) or low (5) price.

5. S accepts or rejects Ω_2. If it accepts, payoffs are $u_B = \frac{2}{3}(12 - \Omega_2)$ and $u_S = \frac{2}{3}(\Omega_1 - r_S)$. If it rejects the offer, $u_B = 0$ and $u_S = 0$.

The game starts with a chance move, with probability $\frac{1}{4}$ that the seller is tough and probability $\frac{3}{4}$ that the seller is soft. The seller is informed of its value for the object, defining its payoffs. The seller's value becomes its private information. The buyer's initial belief that the seller is tough is $\frac{1}{4}$ as given by the probability of the chance move.

In equilibrium, tough sellers always reject offers of 5 and accept offers of 7. Soft sellers always accept any offer in the final round and always accept 7 in the first round. These moves dominate the alternative moves. To specify an equilibrium, I note the buyer's offers in both rounds, the soft seller's response to an offer of 5 in the first round, and the buyer's belief that the seller is hard in the final round. A perfect Bayesian equilibrium of this model is (5,7;a:1), read as (the buyer's first offer, the buyer's final offer; a soft seller's response if the buyer offers 5 in the first round: the buyer's belief that the seller is tough in the final round).

In this equilibrium, the types of sellers separate in their responses to an initial low offer. I prove it is perfect Bayesian. Begin with the buyer's beliefs after the seller's response to the first offer. According to the equilibrium strategies, tough sellers always reject that offer, and soft sellers always accept it. Only tough sellers should be left after the first round; $\bar{\beta}$ (the buyer's updated beliefs) must equal 1:

$$p(\text{S tough}\,|\Omega_1 \text{ rejected}) = \frac{p(\text{S tough})\,p(\Omega_1 \text{ rejected}|\text{S tough})}{p(\text{S tough})\,p(\Omega_1 \text{ rejected}|\text{S tough}) + p(\text{S soft})p(\Omega_1 \text{ rejected}|\text{S soft})}$$

$$= \frac{(\frac{1}{4})(1)}{(\frac{1}{4})(1) + (\frac{3}{4})(0)} = 1$$

Given a belief that the seller is tough, the buyer will always offer high (7) in the final round because tough sellers will not accept a low (5) offer. In the first round, the buyer's expected payoff from offering 5 is $\frac{3}{4}(7) + \frac{1}{4}(3\frac{1}{3}) = \frac{73}{12} = 6\frac{1}{12}$. The buyer's expected payoff from offering 7 in the first round is 5 because all sellers accept such an offer. The buyer prefers an initial low offer

in the first round. If the seller is soft, it receives 5 if it accepts the first low offer and $4\frac{2}{3}$—that is, $\frac{2}{3}(7)$—if it waits for the final offer, so it prefers accepting the first offer.

The buyer uses a low bid to separate the two types of sellers, with the soft sellers taking the first offer. The buyer learns from its first offer because only hard sellers are left in the second round. This is how the players communicate tacitly through their bargaining. The effects of such tacit bargaining can be traced in the buyer's beliefs. Notice that bargaining is inefficient here: both the parties lose value when they go to the second offer.

Exercise 8.9:

a) Find a pooling perfect Bayesian equilibrium of this bargaining game when the initial draw gives the probability of a tough seller to be $\frac{2}{3}$. How does this change in the initial beliefs shift the balance of bargaining leverage between the parties?

b) Find the range of initial beliefs for the buyer that can support the pooling equilibrium. (Hint: Find the belief where the buyer is indifferent between a high and low offer in the first round.)

Exercise 8.10: Consider a case where there are two types of buyer and one type of seller. One type of buyer is soft and places value 12 on the object, call them soft. The other type of buyer is tough and places value 6 on the object. (Notice that tough buyers will never offer 7 for the object.) The seller is soft: it always places value 0 on the object. The players share a common discount factor of $\frac{2}{3}$. Let the initial probability of a tough buyer be $\frac{1}{2}$.

a) Write down the game in time line form.

b) Find the perfect Bayesian equilibrium of the game.

Models of bargaining with incomplete information get more complicated from here. I provide some references for further reading at the end of this chapter. For now, I give a brief review of some typical results of these models. One type of these models has one side making offers and the other side accepting or rejecting them. Typically, the first side is uninformed about the second's reservation level. "Walking offers" typically result in these models. The first actor begins with a low offer and increases it each round until the second player accepts it. The second player accepts any offer better than its reservation level. Another set of models vary the Rubinstein model by adding uncertainty about discount factors and produce similar results. Both types of model assume that a deal is always possible. For example, stipulate that the buyer's reservation

price is known to be \$50 and the seller's reservation price lies between \$20 and \$40. If the players know that it is possible that there is not a deal that both prefer to no deal, agreement is more difficult. Models where both sides can make offers are more difficult. Multiple equilibria become a major problem in these models because both sides must consider how their offers affect the other side's beliefs about the first side's reservation level. In economics, these bargaining models have been applied to strikes. In political science, they have been used to study prewar bargaining and other international negotiations. Now I move onto a different topic and explore reputations and deterrence.

Deterrence and Out-of-Equilibrium Beliefs

Many scholars argue that reputation is an important motivation behind costly acts in international politics. States may act in international crises not for the intrinsic value of the immediate issue, but rather to deter future challenges to their interests by building a reputation for responding to such challenges. For instance, some argue that it was important for the United States to demonstrate its willingness to use force in the wake of the fall of South Vietnam in 1975. According to this argument, President Ford's use of the Marines to free the Mayaguez, a ship seized by Cambodia, demonstrated U.S. willingness to use force. If the United States used force over such a minor issue, the argument ran, surely it would be perceived as ready to use force against a more important challenge, and thus deter such challenges.

But this argument creates a paradox. Deterrence of important threats requires a reputation for the willingness to suffer the high costs of fighting such threats. If a low-cost use of force always succeeds in creating such a reputation, then all types of deterring powers, regardless of the costs they face, want to engage in the low-cost use of force. The long-run benefits justify the short-run costs. But if deterring powers engage in the demonstrative use of force regardless of the costs they face, the demonstration lacks credibility. Threatening powers are not convinced by the demonstration of the willingness to bear costs precisely because *all* deterring powers carry it out.

There is a further twist on this problem of credibility. Action can also be interpreted as a signal of weakness. Those deterring powers who will pay any cost to block a major threat do not need to prove their willingness to fight. Only those powers whose credibility is in doubt need make such demonstrations. Thus fighting to counter minor threats signals weakness, and even deterring powers who face high costs to repel threats will not wish to fight low-cost threats.

This paradox of credibility is a problem of out-of-equilibrium beliefs. What should other nations infer from non-intervention when they expect the deterring power to always intervene? This paradox will not occur in some cases

because some types of deterring powers will fight, others will not, and there are no beliefs off the equilibrium path. But it does occur when the deterring power always or never fights as a demonstration of will. This section presents a simple model, drawn from Nalebuff 1991, that explores this paradox and how different beliefs off the equilibrium path influence behavior on the equilibrium path.

We focus on the actions of a possible deterring power, D. D must choose whether to intervene in an immediate crisis. The stakes of the crisis to D are x; if D intervenes, it gains x immediately. D also pays costs, c, if it intervenes. These costs are private information for D and are drawn from a uniform distribution from 0 to 1. D knows the true value of c, but the threatening power, T, does not.

T's uncertainty creates the possibility of D's developing a reputation by intervening in the immediate crisis. We do not model directly the dynamics of future crises here. Instead, we assume that T's set of beliefs about D's costs for intervention gives D's value for its reputation. If $\bar{c}$ is the mean of T's beliefs about D's costs, $1 - \bar{c}$ is D's reputation, and let $a(1 - \bar{c})$ be D's value for its reputation. The lower the costs that T believes that D faces for intervening, the stronger is D's reputation. Presumably, the lower T believes D's costs to be, the more likely is T to believe that D will intervene in a future crisis, and so the less likely T is to start a crisis in the future. The value to D of future threats relative to the current crisis is given by a, with $a > 0$. The greater a is, the more D values its reputation relative to the current crisis.

If D's reputation did not matter, D would intervene when the benefits of intervention exceeded the costs, $x - c > 0$. Adding reputations complicates the question. D adds the effect of intervention on its reputation to this simple calculation. But we must add both the effect of intervening and the effect of not intervening on D's reputation. If c_i is the mean of D's costs after intervening and c_n is the mean of D's costs after not intervening, then D intervenes when

$$x - c + a(1 - c_i) > a(1 - c_n).$$

Simplifying, we get

$$x + a(c_n - c_i) > c.$$

When the inequality above is changed to an equality, D is indifferent between intervening and not intervening. Call c^* this critical cost. If $c < c^*$, D intervenes; if $c > c^*$, D does not intervene. The greater D's costs, the less likely it intervenes.

If D intervenes for some values of c and not for others, we can calculate D's reputation after intervening and after not intervening. We just update T's beliefs about D's costs, using Bayes's Theorem. The value of c_i is the mean of all the values of c for which D intervenes. This mean is the midpoint of 0 and c^*, $\frac{c^*}{2}$, because c is uniformly distributed. Similarly, c_n is the midpoint of c^* and 1, $(1 + c^*)/2$. We have three equations in three unknowns, c_i, c_n, and c^* as follows:

$$c^* = x + a(c_n - c_i),$$

$$c_i = \frac{c^*}{2},$$

and

$$c_n = \frac{1 + c^*}{2}.$$

From the two equations for c_i and c_n, we can eliminate c^* and solve for their difference: $c_i - c_n = \frac{1}{2}$. Substitute this value back in the equation for c^* in the last paragraph to get $c^* = x + \frac{1}{2}a$. Substitute c^* back into the equations for c_i and c_n: $c_i = \frac{1}{2}x + \frac{1}{4}a$, $c_n = \frac{1}{2} + \frac{1}{2}x + \frac{1}{4}a$.

So far the model fits some of our intuition about building reputations. First, taking costly actions improves one's reputation for being willing to bear costs. The lower D's costs, the more likely it is to intervene, and its reputation improves if it intervenes. Intervention is a noisy signal about D's costs. T cannot determine exactly what D's costs are if D intervenes. But T has learned that those costs are lower than T originally believed. On the other hand, failure to intervene by D reduces its reputation. Second, the greater the value of the future relative to the current crisis, the more likely D is to intervene now; c^* increases as a increases. Finally, the greater the value of the current crisis relative to costs, the more likely D is to intervene; c^* increases as x increases.

This equilibrium requires D to intervene for some costs and not for others. If D always intervenes (or never intervenes), T's beliefs for nonintervention (or intervention) are off the equilibrium path and cannot be calculated by using Bayes's Theorem. The first equation above still describes the incentives that D must satisfy in the initial crisis. One of the equations for T's beliefs also holds, depending on what all Ds do. I look at the case where no D intervenes and leave the case where all Ds intervene to an exercise. I examine the incentives and beliefs in each of these polar cases to show the effects of beliefs off the equilibrium path on equilibrium behavior.

The case where no D intervenes requires T to infer weakness from intervention. Does the model support such an interpretation? Can we find an equilibrium where no D intervenes? If no D intervenes, then $c^* \leq 0$. We can solve for c_n with the third equation above: $c_n = \frac{1}{2}$. We should expect this answer; if no Ds intervene, T cannot learn about D's costs when it observes nonintervention. But the equation for c_i no longer holds because D never intervenes

in this equilibrium and we cannot calculate c_i by using Bayes's Theorem. Instead, we can choose c_i to be any value from 0 to 1 that we like. From the first equation, we have $x + a(\frac{1}{2} - c_i) < 0$ because $c_n = \frac{1}{2}$ and the inequality must be true when $c = 0$. Choose $c_i = 1$, and this inequality becomes $x < \frac{1}{2}a$. These beliefs are extremely optimistic for the threatening power. If T should observe intervention by D, T concludes that D is the weakest type possible. The condition $x < \frac{1}{2}a$ is required for the nonintervention equilibrium. When $x > \frac{1}{2}a$, Ds that face no cost for intervening now (i.e., $c = 0$) prefer intervention now even though such intervention destroys their reputation. They believe that winning the current crisis is more valuable than sustaining their reputation. The smaller x is, the greater the range of beliefs after intervention (c_i) that can sustain the nonintervention equilibrium. For example, $c_i \geq \frac{1}{2}$ supports the nonintervention equilibrium when $x \leq 0$. This case covers situations where no D would intervene in the current crisis if there were no reputational effects. The cost of intervention always exceeds its immediate benefits.

Exercise 8.11: Find a perfect Bayesian equilibrium where all Ds intervene in the current crisis.

a) Find the relationship among x, a, and c_n that must satisfied for all Ds to intervene.

b) Find the relationship necessary for this equilibrium if T draws the more optimistic inference from nonintervention, $c_n = 1$.

c) Are there any values of a and x such that both the nonintervention and the all-intervene equilibria are possible? If so, describe these cases.

Both of the self-fulfilling equilibria are possible. They differ in the nature of the inference that the threatening power draws from behavior that should not occur. Beliefs off the equilibrium path influence what behavior is rational on the equilibrium path. But are all beliefs off the equilibrium path equally valid inferences? The next section discusses what restrictions can plausibly be placed on inferences off the equilibrium path.

An Introduction to Restrictions on Beliefs

The introduction of beliefs alters the idea of backwards induction. Backwards induction directly compares the outcomes that each available move will produce. We compare equilibrium outcomes with possible outcomes off the equilibrium path. But moves can change beliefs, and beliefs affect which moves are optimal. The evaluation of alternative moves is not so simple after adding beliefs. Because we can choose beliefs off the equilibrium path freely, we can choose odd beliefs that lead to odd actions off the equilibrium path. These odd actions then convince other players to choose peculiar actions on the equilibrium path.

Restrictions on beliefs are a way to address this problem. Nash equilibria judge just the optimality of actions along the equilibrium path; actions off the equilibrium path can be chosen freely. Perfect Bayesian equilibria require actions off the equilibrium path to be optimal as well, but we can choose beliefs off the equilibrium path freely. Restrictions on beliefs judge the plausibility of beliefs off the equilibrium path.

Return to the lawsuit signaling game in Figure 8.2. Consider the beliefs that support the equilibrium (B,B;r,a′). This equilibrium is supported by $(p,\frac{1}{2})$ with $p \geq \frac{3}{5}$. But are these beliefs reasonable? They imply that negligent defendants are more likely than nonnegligent ones to defect from the equilibrium by making a small offer. But nonnegligent defendants gain from defection whenever negligent defendants do, and the opposite is not true. We would like to say that these off-the-equilibrium-path beliefs are unreasonable and eliminate them and the equilibrium they support. Recent work in the field has concentrated on the question of what restrictions should be placed on beliefs off the equilibrium path beyond the weak restrictions imposed by perfect Bayesian and sequential equilibrium.

To introduce the ideas behind restrictions on beliefs, I turn to one of the cuter examples in noncooperative game theory, the Beer-Quiche Game (Cho and Kreps 1987). The Game Theory Bar & Grill has only two items on its breakfast menu, beer and quiche. Only two types of people frequent the Game Theory Bar & Grill, surly guys and wimps. Surly guys like beer for breakfast better than quiche; wimps prefer quiche to beer. But both types want to be left in peace to enjoy their breakfast and value that peace more than what they eat.

You, on the other hand, are a troublemaker. You enter the Game Theory Bar & Grill looking for a fight. But you are also a discriminating connoisseur of the sweet science of boxing. Surly guys are tough; wimps are not—that is why they are wimps, after all. You want to fight a wimp but not a surly guy. You enter the bar knowing that more surly guys eat there than wimps do. To be precise, you believe that 90 percent of the patrons of the Game Theory Bar & Grill are surly. But you can observe what a possible target of your aggression has eaten for breakfast. The question is what you can learn about their type from their breakfast.

The extensive-form game in Figure 8.8 specifies this problem. The game begins with a chance move that determines the type of bar patron, surly or wimpy. There is a .9 chance of a surly guy and a .1 chance of a wimp. For ease of presentation, I have labeled this player as S in the upper branch, for Surly

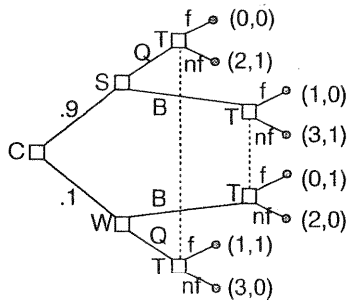

Figure 8.8 The Beer-Quiche Game

Guy, and W in the lower branch, for Wimp. The patron chooses his breakfast, beer or quiche, labeled B or Q. The troublemaker, T, enters the bar and chooses whether to fight the patron (f or nf, for fight or not fight) after observing what the patron ate for breakfast. Payoffs are written (Patron, Troublemaker). The patron receives a payoff of 2 for no fight, compared to 0 for a fight, and an additional 1 for eating his preferred breakfast. For instance, a surly guy who has quiche for breakfast and does not have to fight receives a total payoff of $0 + 2 = 2$. The troublemaker gets 1 for fighting a wimp or not fighting a surly guy and 0 for the opposite outcomes.

One equilibrium of the Beer-Quiche Game is (B,B;f,nf:0,.9) where the equilibrium is read (Surly Guy's breakfast, Wimp's breakfast; Troublemaker's choice after observing quiche, Troublemaker's choice after observing beer: Troublemaker's belief that patron is surly after observing quiche, Troublemaker's belief that patron is surly after observing beer). Wimps imitate surly guys in this equilibrium to avoid a fight. The chance of facing a surly guy is high enough to deter the troublemaker from starting a fight. The wimps are willing to have beer for breakfast because they avoid a fight by doing so. If they eat quiche, they get pounded. Someone eating quiche is off the equilibrium path here. The troublemaker infers that only a wimp would eat quiche when everyone else is having beer for breakfast. The troublemaker, then, prefers fighting anyone eating quiche.

> Exercise 8.12: Demonstrate that (B,B;f,nf:0,.9) is a perfect Bayesian equilibrium of the Beer-Quiche Game.

But there is another perfect Bayesian equilibrium of this game, where surly guys eat quiche: (Q,Q;nf,f:.9,0). The troublemaker is deterred by seeing the patron eat quiche. The chance that the patron is surly is still too high to take the chance on a fight. Here, beer for breakfast is off the equilibrium path. Perfect Bayesian equilibrium allows us freely to choose the troublemaker's beliefs after someone has beer. I choose to let the troublemaker conclude that anyone

having beer for breakfast is a wimp. So the troublemaker is happy to fight anyone having beer for breakfast. Everyone has quiche because peace is more important than the preferred breakfast. Wimps do better in the quiche-eating equilibrium than in the beer-for-breakfast one. They get to eat their quiche in peace. Surly guys eat quiche because beer for breakfast convinces the troublemaker that you are a wimp. Maybe real men do eat quiche.

> Exercise 8.13: Demonstrate that (Q,Q;nf,f:.9,0) is a perfect Bayesian equilibrium of the Beer-Quiche Game.

This second equilibrium seems implausible. In the first equilibrium, wimps paid a price by having beer for breakfast in order to avoid a fight. The troublemaker construes quiche for breakfast as a signal that the patron is a wimp. This interpretation of the patron's behavior is a belief off the equilibrium path because no patron eats quiche in the first equilibrium. In the second equilibrium, the troublemaker understands a breakfast of beer as a signal of wimpiness; if you have beer for breakfast, you are a wimp. This interpretation should strike you as odd. A more plausible inference off the equilibrium path here is that beer signals that the patron is surly. Surly guys prefer beer for breakfast while wimps prefer quiche. If either type defects from this equilibrium, it must be a surly guy. Wimps cannot benefit by having beer for breakfast; their equilibrium payoff is 3, and the most they can receive if they have beer is 2. A surly guy can benefit by defecting if that defection convinces the troublemaker to leave him alone. His payoff would rise from 2 to 3. Why would you infer that a defector is a wimp then? The critical idea here is that wimps cannot benefit by defection from the equilibrium whereas surly guys might.

The opposite is true of the first equilibrium. Wimps might benefit by defection because they prefer quiche to beer. Surly guys cannot benefit even if they eat their quiche in peace. In this case, the inference off the equilibrium path that defectors are wimps is plausible. Wimps have an incentive to defect from the equilibrium of beer for breakfast, and surly guys do not.

This idea can be generalized. Beliefs off the equilibrium path should place zero probability on types that cannot ever benefit by defecting from equilibrium behavior. This restriction unravels the equilibrium where all patrons eat quiche. It requires the troublemaker's belief that the patron is a wimp to be 0 if the patron has beer for breakfast in the second equilibrium. Wimps cannot benefit by having beer for breakfast over their equilibrium payoff. But then not fighting rather than fighting is the optimal response for the troublemaker if the patron has beer for breakfast in this equilibrium. If the probability of facing a wimp is 0, the patron is undoubtedly surly, and the troublemaker does not want to fight a surly guy. Finally, surly guys want beer for breakfast over quiche once the troublemaker will not pick a fight with them after seeing them have beer for breakfast. By restricting beliefs off the equilibrium path to place

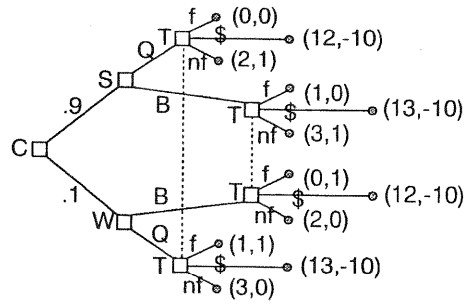

Figure 8.9 The Beer-Quiche Game with Bribes

zero weight on the type that cannot possibly benefit from defecting, we eliminate the (Q,Q;nf,f:.9,0) equilibrium. The threat of fighting patrons who have beer for breakfast is based on an implausible inference about who is likely to defect from this equilibrium.

Although this initial restriction on beliefs is reasonable, there are many equilibria of many games where it provides little ability to eliminate equilibria based on implausible beliefs. It can be strengthened in several ways that I briefly sketch here. My purpose here is not to provide a complete treatment of restrictions on beliefs. The ideas behind these restrictions are intuitive, but the formalisms are not. Further, there is not one commonly accepted set of restrictions on beliefs. An informal treatment of these ideas should help you understand what beliefs off the equilibrium path are eliminated.

First, we broaden this initial criterion for beliefs off the equilibrium path to assess the rationality of both parties in response to a possible defection. As before, consider two types of senders, t and t′. Both types' interest in defecting from an equilibrium depends upon the response of the receiver to that defection. The only responses that could make defection attractive to the sender are out of equilibrium and occur off the equilibrium path. Otherwise, our equilibrium would not be an equilibrium: at least one type of sender would do better by defecting. The second restriction on beliefs assesses the ranges of responses that might provoke defection across types of senders. If type t wants to defect for any response that makes type t′ want to defect, then the receiver's beliefs at the node following that defection cannot increase the probability of type t′ relative to that of type t. The receiver should believe that type t is at least as likely to defect as type t′.

An example may help clarify this idea. Add another response to the Beer-Quiche Game; the troublemaker can give the patron \$10. Figure 8.9 gives the modified game. The dollar sign denotes the move of paying the patron.

This game has the same two equilibria as the Beer-Quiche Game, (B,B;f,nf:0,.9) and (Q,Q;nf,f:.9,0). Fighting and not fighting dominate paying the patron for the troublemaker regardless of her beliefs at either node. The first restriction on beliefs does not bind on the quiche-eating equilibrium of the modified game. There is now a response by the troublemaker, paying the patron, that makes defection better than the wimp's equilibrium payoff. But any response that makes a wimp defect from the quiche-eating equilibrium also makes a surly guy defect. The wimp prefers beer for breakfast when the troublemaker plays mixed strategy (p nf, q f, $1 - p - q$ \$), with probabilities p and q such that $p + q \leq 1$ and $2p + 0q + 12(1 - p - q) \geq 3$. Adding 1 to each side of the inequality, we get $3p + 1q + 13(1 - p - q) \geq 4 > 2$, which implies that surly guys also prefer beer for breakfast against (p nf, q f, $1 - p - q$ \$). The troublemaker's beliefs off the equilibrium path cannot increase the probability of a wimp from its initial beliefs under the second restriction because surly guys defect whenever wimps defect. The troublemaker must believe that there is at least a .9 chance of a surly guy and so no longer prefers to fight if the patron has beer for breakfast. As before, surly guys then want to have beer for breakfast.

Exercise 8.14: Show that the (B,B;f,nf:0,.9) equilibrium of the modified game survives after applying the second restriction on beliefs.

a) Find the inequality that expresses when surly guys wish to defect from the beer-drinking equilibrium.

b) Do wimps also want to defect whenever surly guys do?

c) What beliefs are consistent with the restriction?

d) What is the troublemaker's optimal strategy given these beliefs, and what beliefs support this equilibrium?

e) Why do wimps continue to have beer for breakfast?

The third restriction on beliefs uses the ideas above repeatedly to hone the set of plausible beliefs down. First, ask what range of beliefs meet the second criterion above and update the range of possible beliefs. Second, find the set of optimal responses for the receiver given those beliefs. Third, find which types of senders wish to defect for the responses found in the second set, and apply the second restriction on beliefs again. Repeat these three steps until the set of beliefs and responses does not change.

Return to the modified Beer-Quiche Game and test the plausibility of beliefs in the (B,B;f,nf:0,.9) equilibrium. Let p be the troublemaker's out-of-equilibrium belief that the patron is surly if he has quiche for breakfast. By the second restriction, $p \leq .9$ as shown above. The troublemaker's optimal response is fighting if $p < .5$ and not fighting if $p > .5$ ($p = .5$ produces

indifference). If $p > .5$, wimps defect to quiche, but surly guys do not. Updating p produces $p = 0$. This repetition of the restriction on beliefs eliminates all $p > .5$. The only plausible beliefs have $p \leq .5$, then.

> Exercise 8.15: Apply these restrictions on beliefs to the self-fulfilling equilibria of the geopolitical deterrence and reputation game discussed earlier in this chapter. Begin with the equilibrium where all deterring powers (Ds) intervene.
>
> **a)** What type of D is most likely to defect from the equilibrium by not intervening? Recall that D's type is defined by its costs for intervening, with $0 \leq c \leq 1$, and that the response of the threatening power, T, can be summarized by the reputation c_n.
> **b)** What beliefs can T hold after a defection by D from this equilibrium?
> **c)** Given these beliefs off the equilibrium path, when do all Ds prefer intervention to nonintervention? Discuss; what cases does this represent, and does the restriction on beliefs reduce the set of equilibria?
>
> Now turn to the equilibrium where no D intervenes.
>
> **d)** What type of D is most likely to defect from the equilibrium by intervening?
> **e)** What beliefs can T hold after a defection by D from this equilibrium?
> **f)** Given these beliefs off the equilibrium path, when do all Ds prefer nonintervention to intervention? Discuss; what cases does this represent, and does the restriction on beliefs reduce the set of equilibria?

This section gives you some sense of the basic ideas behind restrictions on beliefs. If you are interested further in this topic, I provide references for further reading at the end of the chapter. The reader should be warned; restrictions on beliefs is a highly technical subject. There are several different ideas of such restrictions, and their differences are subtle and generally expressed technically. I turn now to one final issue in signaling, how costless signals can communicate information.

"Cheap Talk" and Coordination

The examples of signaling games so far use costly signals. The sender's payoffs vary with the signal it sends. Wimps do not like beer for breakfast; acting like a surly guy is costly for them. Costly signals aid separation of types. Some

		Player 2 s_1	Player 2 s_2
Player 1	S_1	(a,1)	(0,0)
	S_2	(0,0)	(1,a)

Figure 8.10 General Battle of the Sexes

types bear a cost for attempting to imitate other types, and sometimes those costs outweigh the benefits of imitation. This section covers costless signaling, better known as "cheap talk," from the saying "Talk is cheap." It addresses how communication can help players select among multiple equilibria of games with complete information. Unlike the examples earlier in this chapter, this section does not present a game of limited information.

Cheap talk can be effective in fostering coordination of strategies among players. Players often share an interest in coordinating their strategies for mutual benefit. For example, consider the problem of deciding on which side of the road to drive. If no one owns a car, driving on the right side or the left side of the road are equally attractive. What matters is that we all agree on which side we will drive. Talking among ourselves or coordinating around an external signal should suffice to get us to all drive on the same side of the road. The problem is more complicated if some of us own cars with the steering wheel on the left, while others own cars with the steering wheel on the right. We continue to share an interest in agreeing to drive on the same side of the road, but those with steering wheels on the left prefer an agreement to drive on the right side of the road to an agreement to drive on the left side (and vice versa for those with steering wheels on the right). Can costless signals help us to reach an agreement about the side we will drive on in the face of these divergent preferences?

The Battle of the Sexes game, introduced in Chapter Four, is the simplest representation of this problem. Shown in Figure 8.10, Battle of the Sexes captures a situation where the players wish to coordinate their strategies but disagree on which strategy they prefer to coordinate. We assume that $a > 1$; Player 1 prefers that they play $(S_1; S_1)$, and Player 2 prefers that they play $(S_2; S_2)$. Both prefer coordinating on the other's preferred strategy to not coordinating.

Battle of the Sexes has three equilibria: $(S_1; s_1), (S_2; s_2)$, and

$$\left[\left(\frac{a}{a+1}S_1, \frac{1}{a+1}S_2\right); \left(\frac{1}{a+1}S_1, \frac{a}{a+1}S_2\right)\right],$$

which provide values of (a,1), (1,a), and

$$\left(\frac{a}{a+1}, \frac{a}{a+1}\right)$$

to the players. Battle of the Sexes is a symmetric game because the players are identical. If we believe that symmetric games should have symmetric equilibria, then the mixed strategy equilibrium is chosen equilibrium. The argument here is simple. The pure strategy equilibria require both players to know which strategy to play in advance and have sufficient confidence that the other player will play that same strategy. But from where does such a common conjecture arise? If the players do not have the opportunity to communicate with one another before selecting their strategies, such a common conjecture seems highly unlikely.

Cheap talk can help create shared expectations between players. Assume that players can exchange messages before selecting their strategies. Further, restrict the set of possible messages to the pair "I want to play $(S_1;s_1)$" and "I want to play $(S_2;s_2)$." Call these messages $\underline{1}$ and $\underline{2}$ respectively. The addition of these costless messages creates a game with the following time line:

1. The players simultaneously exchange messages from the set $(\underline{1},\underline{2})$. Both messages are revealed.
2. Both players simultaneously choose their moves in the Battle of the Sexes.
3. Moves are revealed, and payoffs received.

An equilibrium of this game requires that the players select optimal moves given the messages they have received and that they select optimal messages given their expectations about what moves will result from those messages. What equilibria does this game have?

The first equilibrium is depressing: the players do not use their messages to coordinate their moves. Player 1 always sends $\underline{1}$, and Player 2 always sends $\underline{2}$ (or vice versa), and both play the mixed strategy equilibrium of Battle of the Sexes for their moves regardless of what message they have sent or received. Each player anticipates that the other will use its mixed strategy to determine its move, so using its own mixed strategy to choose its move is a best reply. The signal sent has no effect on the other player's choice of move, so any signal is optimal. The important feature of this **babbling equilibrium**, as it is called, is that neither player conditions its move on the signals. The signals "babble" about the moves that the sending player intends; they provide no clue about what move that player will select. Neither player can condition its move on the signals.

All games with cheap talk have babbling equilibria. It is always possible to find an equilibrium where the players do not use the signals to coordinate their moves later in the game. It is more interesting to ask if they can use the signals to help them coordinate their moves.

The second equilibrium uses a straightforward interpretation of the signals. If the players' signals agree, they both play that strategy. If their signals disagree, they play the mixed strategy equilibrium of Battle of the Sexes. If both players understand that they will both use this rule in their choices of both signals and moves, what is the equilibrium? Begin with the selection of moves, and focus on Player 1 for convenience (Player 2's choices mirror these calculations, through symmetry). If $(\underline{1},\underline{1})$ are the messages—read this as (Player 1's message, Player 2's message)—then Player 1 should play S_1 because he expects that Player 2 will play s_1. To do otherwise reduces his payoff. Similarly, if $(\underline{2},\underline{2})$ are the messages, then Player 1 should play S_2 because he expects that Player 2 will play s_2. If $(\underline{1},\underline{2})$ or $(\underline{2},\underline{1})$ are the messages, Player 1 expects that Player 2 will play her mixed strategy for Battle of the Sexes. His mixed strategy then is a best reply to her mixed strategy. The strategies then are optimal given the signals.

But what signal should Player 1 send? Let p be the probability that Player 2 sends $\underline{1}$. The probability that she sends $\underline{2}$ is $1 - p$. If Player 1 sends $\underline{1}$, there is p probability that Player 2 also sends $\underline{1}$, and they coordinate on $(S_1;s_1)$, which gives Player 1 a payoff of a. There is $1 - p$ probability that Player 2 sends $\underline{2}$, leading to the mixed strategy equilibrium of Battle of the Sexes. Player 1 receives an expected payoff of $a/(a + 1)$ in that case. If Player 1 sends $\underline{2}$, there is $1 - p$ probability that Player 2 also sends $\underline{2}$, and they coordinate on $(S_2;s_2)$, which gives Player 1 a payoff of 1. There is p probability that Player 2 sends $\underline{1}$, leading to the mixed strategy equilibrium of Battle of the Sexes and an expected payoff of $a/(a + 1)$ for Player 1. Find p for Player 2 that makes Player 1 indifferent between sending $\underline{1}$ and $\underline{2}$. Calculate his expected utility for sending each signal and set them equal:

$$p(a) + (1 - p)\left(\frac{a}{a+1}\right) = p\left(\frac{a}{a+1}\right) + (1 - p)1$$

$$p = \frac{1}{a^2 + 1}.$$

Player 2's mixed signaling strategy that makes Player 1 indifferent between sending $\underline{1}$ and $\underline{2}$ is

$$\left(\frac{1}{a^2+1}\underline{1}, \frac{a^2}{a^2+1}\underline{2}\right).$$

By symmetry, Player 1's optimal signaling strategy is

$$\left(\frac{a^2}{a^2+1}\underline{1}, \frac{1}{a^2+1}\underline{2}\right).$$

Both players send a signal that they are willing to coordinate on their less preferred strategy pair—which is $(S_2;s_2)$ for Player 1—occasionally because coordination on that pair is preferable to playing the mixed strategy equilibrium.

This cheap talk equilibrium makes both players better off than the mixed strategy equilibrium does. In the mixed strategy equilibrium, there is a probability of

$$\frac{2a}{(a+1)^2}$$

that the players successfully coordinate their moves. Each player expects $a/(a+1)$ from the mixed strategy equilibrium. In the cheap talk equilibrium, the players successfully coordinate their moves with a probability of

$$\frac{2a(a^2+a+1)}{(a+1)^2(a^2+1)}.$$

This chance is greater than the probability of successful coordination in the mixed strategy equilibrium by a factor of

$$\frac{a^2+a+1}{a^2+1} > 1.$$

Each player's expected payoff from the cheap talk equilibrium is

$$\frac{a(a^2+a+1)}{(a+1)(a^2+1)}.$$

This payoff is greater than its expected payoff from the mixed strategy equilibrium by a factor of

$$\frac{a^2+a+1}{a^2+1} > 1.$$

Both players are better off because they receive the mixed strategy payoff when their signals do not match and a better payoff when the signals do match. The improvement produced by cheap talk declines as the distributive issue in the underlying Battle of the Sexes increases. The magnitude of a reflects the distributive problem here: the greater a is, the greater risks each player accepts to gain its preferred strategy pair. The greater a is, the smaller the increases in probability of successful coordination and payoff that cheap talk can produce in this equilibrium. Both factors approach 1 as a grows without bound.

Exercise 8.16: We can improve the performance of cheap talk by allowing the players to exchange two rounds of messages. Again restrict the messages to the set $\underline{1}$ and $\underline{2}$, and keep the interpretation of the second-round messages the same. Change the interpretation of the first-round messages to be as follows: If the signals match, play that move and do not send a second signal. If they do not match, send a second signal and follow the cheap talk equilibrium above. Then find Player 1's optimal mixed strategy for his first-round signal. Is the probability of Player 1's sending a signal of $\underline{2}$ higher in the first or second round? (Hint: If the players have to send a second signal, they will follow the cheap talk equilibrium above for those signals. Those with a taste for messy algebra can calculate the probability that the players coordinate their moves and their payoffs, using two rounds of cheap talk.)

Two rounds of cheap talk improves the efficiency of cheap talk. Coordination is more likely, and the players' expected payoffs increase over one round of cheap talk. But there still remains a significant probability that the players will fail to coordinate their moves.

There is another form of cheap talk that achieves complete efficiency by always coordinating the players' moves. Change the interpretation of the signals as follows: If the players' signals match, that is, $(\underline{1},\underline{1})$ or $(\underline{2},\underline{2})$, play $(S_1;s_1)$; if they differ, play $(S_2;s_2)$. Given that the players expect each other to follow this rule, it is optimal to play the moves dictated by each pair of signals. If Player 1 believes that Player 2 will play s_2, then he should play S_2. What are the optimal signaling strategies? Let p be the probability that Player 2 sends $\underline{1}$ and choose p to make Player 1 indifferent between sending $\underline{1}$ and $\underline{2}$:

$$a(p) + 1(1-p) = 1(p) + a(1-p)$$

$$p = \tfrac{1}{2}.$$

Player 1 should send the signals with equal probability. The players always coordinate their moves in this equilibrium and receive expected payoff $(a+1)/2$ in it. I call this equilibrium a **matching pennies** equilibrium because the logic of the signals is identical to that of the Matching Pennies game in Chapter Three. Each player wants its preferred equilibrium to be chosen, but both signals together determine which equilibrium they play. My optimal signal is obvious if I know what signal you are sending. Thus you must make your signal unpredictable to protect yourself.

There are other ways to achieve perfect efficiency besides the matching pennies equilibrium. Both of the asymmetric equilibria of Battle of the Sexes also achieve perfect equilibrium. But they create a distributional problem between the players. If the game is repeated several times, the players can agree to alternate between the two equilibria. Alternation does create a distributional problem if the players discount future payoffs. The player who first receives

its preferred pair of moves benefits because its large payoff is discounted less than that of the player who receives its big payoff second. The players could also use external events to coordinate their moves. They could flip a coin and play $(S_1;s_1)$ if the coin came up tails and $(S_2;s_2)$ if it came up heads. This agreement makes both players better off than the mixed strategy equilibrium ex ante. All of these agreements require the players to communicate their shared understanding about how they will coordinate their moves.

This example of cheap talk may seem trite and excessively abstract to you. But cheap talk is a way to address a number of very important questions about politics. For example, legislative debate is basically cheap talk.[4] Debate provides a way for legislators with similar underlying preferences to coordinate their votes. Speeches by knowledgeable members serve as cues about how less-informed members should vote on complex legislation. Cheap talk is most effective when the sender and the receiver hold similar preferences over the outcomes. Members are unlikely to take cues from those whose underlying values are greatly different from their own. I provide a brief guide to the literature on debate as cheap talk in the references. Cheap talk also provides a way to think about some issues in negotiations. Battle of the Sexes is a simple representation of a bargaining problem. Both sides wish to reach an agreement, but they differ about the exact form of the agreement. If there is a cost to delaying agreement, then the players can speed an agreement though cheap talk. I have used this idea to explore bargaining under international regimes (Morrow 1994a). Regular rules of bargaining assist the parties in exchanging information that can help them reach mutually beneficial agreements. Nations almost always have incentives to misrepresent their position in negotiations. Common understandings of how disputes will be negotiated can assist in solving those disputes.

Review

This chapter has covered games under limited information, focusing on signaling games. These games provide a way to explore important questions in politics that cannot be addressed by using classical game theory. Limited information creates the possibilities of misperception, deception, and honest communication among the players. Signaling games are now a standard way to analyze questions where all of these issues are central.

A game of limited information is modeled by a set of games, one game for each possible combination of the types of the players. A chance move determines which combination of types will actually be played. Each player knows only its own type. The players' beliefs express the likelihood of the different types of other players in the game. Because a player's type can influence its moves, the other players can infer a player's type from its moves.

The strength of these inferences depends on equilibrium behavior. In a pooling equilibrium, all types of a player take the same action, so other players cannot infer anything new about its type from that action. In a separating equilibrium, each different type of a player takes a distinct action, so the other players can determine its type by observing its action. A separating equilibrium requires incentive compatibility. No type can prefer taking the action of another type to its own action in equilibrium.

Pooling equilibria raise the question of beliefs off the equilibrium path. When all types take the same action, any beliefs after observing any other action are off the equilibrium path. Strange behavior on the equilibrium path can be driven by strange beliefs off the equilibrium path. In some situations, we can judge some beliefs off the equilibrium path as unreasonable. The general principle is that beliefs off the equilibrium path should not increase the probability of types that have less incentive to defect than other types.

Further Reading

Most of the literature here is quite difficult to read. The original approach was presented in Harsanyi 1967–1968. However, this topic only took off with the sequential equilibrium concept and the addition of beliefs.

The signaling game presented here is lifted from Banks and Sobel 1987. Banks 1991 is an excellent technical introduction and survey of how signaling games have been applied to numerous areas of political science.

The informational approach to congressional structure draws on research by Thomas Gilligan and Keith Krehbiel. The model in the text is taken from Gilligan and Krehbiel 1987. My presentation of their model has been helped by the clear presentation in Epstein and O'Halloran 1993. Gilligan and Krehbiel (1989, 1990) advance their work on the informational role of congressional committees. Krehbiel (1991) presents this view informally.

The best beginning source on noncooperative bargaining theory is Kennan and Wilson 1993. I have also found Sutton 1986 and Wilson 1985 useful. Binmore 1992, Fudenberg and Tirole 1991, Kreps 1990a, Myerson 1991, and Rasmusen 1989 all have chapters on bargaining.

The model on deterrence and out-of-equilibrium beliefs is taken directly from Nalebuff 1991. Nalebuff provides a good discussion of the different ideas of beliefs off the equilibrium path and how they apply to his model.

If you wish to read more about restrictions on beliefs, I recommend one of the textbooks. Look for their sections on equilibrium refinements. Chapter Eleven of Fudenberg and Tirole 1991 is the most complete discussion in the textbooks. Original sources for restrictions on beliefs are Banks and Sobel 1987, Cho and Kreps 1987, and Kohlberg and Mertens 1986. Cho and Kreps 1987 is the source of the Beer-Quiche Game.

The term *cheap talk* was coined by Joseph Farrell. His work on cheap talk is central to the topic (Farrell 1987; Farrell and Gibbons 1989). David Austen-Smith has done extensive work applying cheap talk models to legislative debate. See Austen-Smith 1990 and Austen-Smith and Riker 1987, the latter as corrected by Austen-Smith and Riker 1990. Johnson 1993 connects cheap talk in game theory to the idea of free and open debate in critical theory. I apply a cheap talk model to international cooperation in Morrow 1994a.

International Politics

International politics has been a major area for game-theoretic work in political science. Because the number of actors is small, strategic interaction is important in international politics.

Crises and war have been the primary focus of formal models in international politics. O'Neill (1992) provides a comprehensive survey of papers using game theory to analyze issues of war and peace. Models of crisis bargaining draw on the literature on bargaining under incomplete information discussed in this chapter. This literature is now quite large. I recommend that you begin with Kilgour and Zagare 1991; the model in this paper is simple and easy to follow. Morrow 1989 is the first paper in which these models were used to analyze conventional crises. Selection effects in observed crises and how they affect empirical analysis is that paper's central point. Fearon (1994) is clearer about selection effects than Morrow (1989), and he extensively discusses evidence supporting the selection effects hypothesis. Bueno de Mesquita and Lalman (1992) use their crisis model to examine structural theories of international politics, including why democracies do not fight other democracies. Kilgour (1991) also addresses why democracies do not fight one another. Other sources on crisis bargaining are Banks 1990, Langlois 1991, Morrow 1992, and Nalebuff 1991. Banks (1990) proves a general theorem about how success in crises and the probability of war must rise as the challenger's resolve increases. Langlois (1991) provides a general model of punishments that stabilize crises. Morrow (1992) discusses the credibility of linkage as a signal of resolve. Nalebuff 1991 is the source of the model on deterrence and out-of-equilibrium beliefs in this chapter.

Models of nuclear strategy are closely related to models of crisis bargaining. Powell (1990) presents a series of different models of arguments in nuclear crises, including the fear of reciprocal surprise attack, the threat that leaves something to chance, and limited retaliation. Wagner (1991) also discusses the rationality of counterforce strategies.

In the area of security studies, Brams and Kilgour (1988) present a series of models, beginning with their model of a nuclear crisis. Downs and Rocke (1990) investigate how arms control agreements can be enforced in the face of verification difficulties. Wittman (1989) uses a model of limited information to address verification and arms control. Kilgour and Brams (1992) study the stability of arms control agreements. Powell (1993) builds a noncooperative model of arms competition between two nations. Morrow (1994b) models alliance formation for deterrence of a common threat.

In the area of structural theories of international politics, Niou and Ordeshook's (1989, 1990) work on the balance of power is important. Niou and Ordeshook (1989) uses a cooperative game framework to analyze the stability of balance-of-power systems. Niou and Ordeshook (1990) extend their results to a noncooperative game. Wagner (1986) begins to develop a formal model of the balance-of-power theory. Kim and Morrow (1992) address power transition theory by asking when power transitions should lead to war.

In the area of international political economy, Alt, Calvert, and Humes (1988) use a modified Chain Store Paradox (presented in Chapter Nine) to address hegemonic stability theory. Martin 1993 is an informal presentation of a noncooperative model of economic sanctions. Powell 1991 is a simple model of trade and conflict between two nations. Morrow (1994a) uses the notion of "cheap talk" as presented in Chapter Eight to model why the forms of cooperation might vary across issues. Putnam's (1988) informal argument began the formal analysis of the links between domestic politics and foreign policy. Iida 1993 and Morrow 1991 use the framework of the two-level game to address this question. Domestic politics and foreign policy is a central issue in Bueno de Mesquita and Lalman 1992.