norms that are not (perhaps because they are thought to be "cultural" or "political") is impossible to draw with precision.

The only approach, then, is a pragmatic one. The book will address conventional tools of international law—treaties and customary international law, as well as the related soft law and norms—but will not attempt to offer an expansive theory of all forces that impact state behavior. I recognize that there is no clear distinction between the set of materials included in the book and the set that is outside its scope, but the guidelines provided here are sufficient to move forward in a fruitful way as long as we remain aware that other forces are also at work.

Ultimately, the goal of this book is to advance our understanding of international law and to do so with conventional social science methodology. I believe that the rational choice assumptions made in the book are appropriate for that exercise. The challenge of this book is to advance a coherent and general theory of how international law influences state behavior. It is my hope that the discussion and analysis that follows will be of interest to both lawyers and social scientists interested in international law and, indeed, to anyone interested in the international legal system.

2

A GENERAL THEORY OF INTERNATIONAL LAW

International law comes in many different guises. It includes treaties, soft law, customary international law, general principles of law, and perhaps more. Though there are plenty of theories of international law that attempt to explain one or more of these categories, there remains a need for a theory that explains state behavior across the full spectrum of international law and legal instruments. It is the task of this chapter to construct just such a theory. The goal is to demonstrate how international legal obligations might influence state behavior, when international obligations are likely—or unlikely—to make a difference, and to begin our exploration—continued in later chapters—of how this theory applies across the various areas and sources of international law.

Games States Play

The subsection that follows describes games in which states find it relatively easy to cooperate, even under our rational choice assumptions. These situations provide only the most modest test for the relevance of international law. The subsequent section turns to the prisoner's dilemma, in which cooperation is much more difficult. It is in the context of this latter game that the theory is applied throughout most of the book.

Simple Forms of Cooperation

Even in a world of selfish states, there will be times when states have common interests that make cooperation valuable and easy to achieve. Consider, for example, the relationship between Canada and the United

States. For a variety of reasons, neither state stands to gain from a military confrontation. Canada would obviously like to avoid such a conflict because it could not hope to emerge victorious, while the United States, despite its military superiority, has no interest in using force against Canada. The ongoing relationship between the two states is worth more to the United States than what would emerge from the use of force. Canada and the United States could choose to sign an agreement explicitly reaffirming the legal obligation not to attack one another, but it is not clear what such a treaty would achieve.[1] They are likely to behave in the same way with or without a treaty because of their shared interest in avoiding a military conflict. Though an international agreement may be put in place in such a situation, it has little role to play beyond perhaps generating some good feeling between the states and their citizens and providing a cheap signal of good intentions that is in line with both parties' prior expectations.

When treaties or other international agreements exist in such situations, one would expect a high rate of compliance, but one would not conclude that international law is achieving anything. An agreement of this kind is akin to a treaty stating that the citizens of treaty signatories will breathe in oxygen and breathe out carbon dioxide. The treaty is in some sense complied with, but it does not do any work.[2] These games lead to cooperation as a result of the underlying payoffs (e.g., both Canada and the United States enjoy higher payoffs if they refrain from military action against one another) rather than anything the international agreement does. An international agreement that imposes an obligation on each party to behave as it would do anyway does not advance international cooperation in any meaningful way, and can hardly be described as effective international law.

International law does only slightly more work in pure coordination games. These are games in which all players have an incentive to cooperate, but cooperation requires that they coordinate their actions. One classic example of such cooperation is the system of rules and regulations governing international air travel and air safety. A series of international agreements, beginning with the Convention for the Unification of Certain Rules Relating to International Carriage by Air (the "Warsaw Convention"), operate to harmonize a range of standards.[3] This represents a successful resolution of the coordination game that air travel presents to regulators. In this and other coordination games, the players in the game (i.e., states) have an interest in coordination that

trumps their interest in a particular outcome. The benefits of coordination in the regulation of air travel are obvious. If airplanes in France were subject to one set of safety regulations but those in the United States another, the cost of travel (especially because a single flight might cross over many countries) would rise dramatically. Even for mundane matters such as the tagging of baggage, coordination makes the operation of air travel a great deal more efficient and its regulation more effective.

One might wonder why states bother with formal agreements to address coordination problems of this sort when simpler strategies would achieve the same result at lower cost. States could (and do), for example, use very simple forms of international communication such as a joint memorandum or an exchange of letters to elicit compliance. Research in economics has suggested that such cooperative methods of coordination can be superior to noncooperative, "market leadership" methods of generating coordination (Farrell and Saloner 1988). Though these cooperative measures could be described as forms of agreement, we will not tarry over that semantic question.

In most situations, states do not use formal agreements to resolve coordination problems. Thus, an agreement for two heads of state to meet will involve reliance and the expenditure of resources by both sides but will not typically be the product of a formal treaty or significant international agreement. The date will be agreed to, and officials from both states will exchange information of all sorts, but no international agreement of note will be produced.

In other cases, however, states do use more formal agreements. One explanation for why they choose to do so is that interactions that appear to be coordination games may in fact be some other type of game. States have an interest in certainty and predictability over time and may, therefore, want an agreement that will offer some assurance about how others will behave if the payoffs change at some future date. In other words, once we consider the potential for future developments, the game may not be a pure coordination game, and states seeking some assurance of future cooperation may want a more formal agreement to encourage such cooperation. I discuss this possibility in greater detail later in the chapter. An alternative and consistent explanation is that the creation of a formal instrument to facilitate coordination costs very little. For example, NAFTA is primarily an agreement about trade and investment, but includes a provision stating that the Free Trade Commission,

which consists of cabinet-level representatives from each party to the agreement, will meet at least once a year and that each party successively will chair the sessions of the Commission. This particular provision is just a way to ensure that meetings take place regularly. The same result could almost certainly be achieved through an informal agreement, but since the states were already negotiating NAFTA, it was convenient and inexpensive to address the coordination problem of when and where to meet in the same document. Finally, states may use a formal agreement to address a coordination game if there is uncertainty with respect to payoffs. If a state is unsure about the payoffs facing another state, it may be concerned that what appears to be a coordination game may in fact be a more difficult-to-solve cooperation problem. Entering into a formal agreement may help to guard against that possibility.

Coordination games can be distinguished from each other by the degree of tension that exists between the interests of the parties (Ginsburg and McAdams 2004). In the so-called battle of the sexes game, for example, the parties face a distributional conflict. They both strictly prefer coordinating their actions to not coordinating, but the players prefer to coordinate on different equilibria. After a focal point (meaning a solution that seems natural or relevant for the parties) is chosen, the parties have no incentive to defect, but the process of choosing a specific outcome may be sensitive to differences in bargaining power among the parties (Krasner 1991). The allocation of radio frequencies and policies addressing satellite communication is arguably an example of such a situation (Krasner 1991). Other examples of coordination games are easy to find. For example, trains running from Spain to the rest of Europe must pass through France, yet historically Spanish rail gauges were wider than the international standard rail gauges used by France. The result is that trains traveling on the broad-gauge Spanish railways must pass through gauge-change installations when crossing the border. To address this inefficiency, new high-speed trains and rails connecting Spain to France and the rest of Europe have been built using the international standard-gauge width.

Consider one further example, the hosting of the Olympic Games. States that would like to host the games in a particular year have conflicting interests. If Paris hosts the games, New York will not be able to do so. Cooperation is more difficult than in a pure coordination game, because the United States would like the games to be in New York while France would like them to be in Paris. Until a location is chosen, then,

the parties' interests are, to some extent, divergent. Once a host city is chosen, however, neither state has an incentive to defect. If New York is chosen for the Olympics, France is better off sending its athletes to New York than boycotting the games or attempting to stage some competing set of games in Paris.[4]

Once again, although states sometimes use international law to resolve such games, there are many other ways to achieve the same goal. It can be done through unilateral actions, repeated practice, informal or tacit agreements, third-party intervention, and so on. In the case of the Olympics, the International Olympic Committee, which is a nongovernmental organization (NGO) over which states have no direct control, determines the host city and state. Once that decision is announced, no state—including one that wanted but failed to be chosen as host—has an incentive to defect. Cooperation is achieved without any need for enforcement, any international legal commitment, or even any action by states.

This set of three games (common interest, pure coordination, and battle of the sexes) does not exhaust the list of games in which cooperation is straightforward and enforcement is not required.[5] It is, nevertheless, sufficient to illustrate the point that in some circumstances cooperation is easy. In such situations, international law does no heavy lifting and might even be said to be superfluous. It may be used to facilitate cooperation, but even then its role is modest. International law is only worthy of study if it does more.

Notice that one would expect a high level of compliance with international law if a large share of international law commitments existed to solve coordination games. This illustrates why, in addition to the possibility that international law often calls on states to do what they would have done anyway (Downs, Rocke, and Barsoom 1996), the observed high rate of compliance (Henkin 1979) may not indicate that international law has a significant effect on state behavior.

Difficult Cooperation: Prisoner's Dilemma

Many international problems are more difficult to solve than those represented by the aforementioned games. It is these more challenging problems that are of primary interest in this book. The greatest value of international law, after all, is its ability to facilitate cooperation in an anarchic world, and this is especially true when cooperation is difficult

to achieve in any other way. Focusing on situations where cooperation is difficult is also helpful as an analytic matter. If international law is able to generate cooperation in these situations, it is because it is having a real impact on behavior.

The bilateral prisoner's dilemma is used to get a sense of how international law can deal with these more difficult problems of cooperation. Similar issues and similar problems of cooperation arise in multilateral contexts. Multilateral cooperation also generates some additional challenges that I will address later in this chapter.

As will be familiar to many readers, a prisoner's dilemma is a game in which the parties can maximize their total joint payoff through mutual cooperation but each player does better by defecting. The familiar result is that in a one-shot game the only equilibrium is for both parties to defect, leading to a low payoff for both players.

Bilateral arms control negotiations provide a clear example of a prisoner's dilemma at work in an international law context. Though any number of arms agreements could be considered, the focus here is on the negotiations between the United States and the Soviet Union leading up to the 1972 Anti-Ballistic Missile (ABM) Treaty. Though the United States and the Soviet Union had many conflicting interests, they had a shared interest in curtailing the nuclear arms race. A series of arms control agreements, both multilateral and bilateral, (the Strategic Arms Limitation Treaties, known as SALT I and SALT II; the Non-Proliferation Treaty (NPT); the Limited Test Ban Treaty) were negotiated over the course of the 1960s and 1970s. The ABM Treaty was aimed at solving a special problem within the broad area of arms control. The acquisition of weapons and technology by both sides represented a significant drain on the resources of these two states. As long as both states were doing the same thing, it did not deliver an increase in security to either one. Furthermore, if one side developed an ABM system capable of protecting against ballistic missiles, the other side would have to invest heavily in developing its own ABM system or missiles that could penetrate the opposition's ABM system. A defensive ABM system, therefore, had the potential to greatly accelerate the arms race. While both sides stood to gain from arms control, each also had much to gain from unilateral defection from an agreement. If one side could continue with its arms program while the other side did not, the former faced the potentially enormous payoff that would come to whichever side developed the ability to wipe out the other's second strike

capability. The problem, then, was a classic prisoner's dilemma. The best collective outcome was mutual cooperation (arms control), but the dominant strategy for each side was to cheat (arms race). The ABM Treaty represented an attempt to overcome this prisoner's dilemma and allow the states to achieve the first best outcome of mutual arms control.

To illustrate how cooperation can come about in the ABM context as well as in other prisoner's dilemmas, imagine a simplified version of the payoffs facing the United States and the Soviet Union. If both sides lived up to their commitments, they would each enjoy a relatively high payoff of, say, 100, representing a world in which the economic strain from the arms race was reduced. If both sides violated the agreement, neither would gain a strategic advantage over the other, but both would bear the heavy economic burden of a continued arms race. This behavior would lead to a lower payoff of, say, 80. If the United States complied with the agreement while the Soviet Union violated it, the United States would suffer a strategic loss of power relative to the Soviet Union—precisely the outcome the United States sought to avoid by participating in the arms race. The Soviet Union, on the other hand, would benefit from the strategic advantage it could gain by violating the agreement while the United States complied. The payoffs from this outcome, relative to the case of mutual compliance, would be very high for the Soviet Union and very low for the United States. We can represent this outcome with payoffs of 200 and −50 for the United States and Soviet Union, respectively. Finally, if the United States violated while the Soviet Union complied with the agreement, the payoffs would be 200 and −50, respectively.[6] These payoffs are represented in table 1.

This is a classic prisoner's dilemma. Though mutual cooperation yields the highest total payoff, each party has an incentive to violate the agreement, regardless of what the other party does.

Table 1. The ABM Treaty Prisoner's Dilemma

		Soviet Union	
		Comply	*Violate*
United States	Comply	(100, 100)	(−50, 200)
	Violate	(200, −50)	(80, 80)

The well-known result for a single play of this game is that both sides will choose to violate and cooperation will fail. Nor can an international law agreement hope to affect this outcome in a one-shot game. An agreement is simply an exchange of promises. Without a system of courts or police capable of enforcing the rights of the parties to the ABM agreement, this exchange of promises has no impact on the payoffs and, therefore, no impact on behavior.

Under these assumptions, if an agreement were put into place, we would expect to see routine violations. Any compliance that might emerge must be attributed to something other than the agreement. Indeed, because the agreement serves no purpose, one would not expect rational states to enter into it at all.

The actual course of events, however, was quite different from the foregoing prediction. The ABM treaty was not only entered into, it was honored by both sides for some time. In fact, the degree of cooperation that it fostered between bitter enemies is impossible to reconcile with the simple one-shot prisoner's dilemma. Something else must have been going on.

That something else was the repeated nature of the game being played. Both states had an interest in future negotiations, and both had an interest in continued compliance by the other side. Violating the agreement would have compromised the agreement reached under the ABM treaty as well as the ability of the violator to make promises about future behavior. It would, therefore, undermine cooperation with respect to both this agreement and future negotiations.

The repeated nature of the interactions between the United States and the Soviet Union allowed cooperation to take place. Each country valued cooperation not only today but also in the future. This gave the parties at least three reasons to comply with the treaty. First, and perhaps most important, is reciprocity. A violation by one side would likely provoke a violation by the other side. The one to violate initially would enjoy a one-period gain, but thereafter the treaty might collapse, in which case both parties would return to the noncooperative outcome. Second, both parties wanted to be able to make credible commitments in the future. By complying with its promises, each country enhanced its reputation as a state that honors its commitments and, therefore, its ability to make future promises. Third, a violation had the potential to trigger some form of retaliatory action, which might further increase the cost of breach.

These three costs—reputation, reciprocity, and retaliation, which I refer to as the Three Rs of Compliance—are the keys to understanding why states comply with international obligations.

The Three Rs of Compliance

Reputation can be defined as judgments about an actor's past behavior used to predict future behavior (Miller 2003). In this book I am interested primarily in a state's reputation for compliance with international law rather than the other types of reputation one can imagine (political scientists, for example, often focus on a reputation for resolve or toughness; Huth 1997; Mercer 1996; Nalebuff 1991). A state's reputation for compliance with legal obligations consists of judgments about the state's past behavior and predictions made about future compliance based on that behavior. Unless some other meaning is specified, I will use the term "reputation" throughout the book to refer to this particular type of reputation. I discuss the relationship between a reputation for compliance with international law and other types of reputation in the next chapter.

The term "reputational sanction" refers to the cost imposed on a state when its reputation is damaged. Reputational sanctions, then, are not punishments at all, or at least they are not intended as such. When a state makes a compliance decision (i.e., when it chooses to comply or violate) it sends a signal about its willingness to honor its international legal obligations. Other states use the information in this decision to adjust their own behavior. A state that tends to comply with its obligations will develop a good reputation for compliance, while a state that often violates obligations will have a bad reputation. A good reputation is valuable because it makes promises more credible and, therefore, makes future cooperation both easier and less costly.

"Reciprocity" refers to actions that, like reputation, will often be taken without the intent to sanction a violator. In response to a violation, states may withdraw their own compliance with an international agreement because once the violation takes place the agreement ceases to serve their interests. A reciprocal action is not costly to the reciprocating state. It is instead an adjustment in a state's behavior motivated by a desire to maximize the state's payoffs in light of new circumstances or information.[7]

"Retaliation," in contrast, describes actions that are costly to the retaliating state and intended to punish the violating party. Retaliatory actions might include, for example, economic, diplomatic, or even military sanctions. Though retaliation and reciprocity often play an important role in encouraging compliance, reputational arguments are required to understand these behaviors. For this reason, I discuss them first.

Reputation

The use of reputational arguments to explain state behavior is familiar in both political science (Axelrod 1984; Keohane 1984) and economics (Abreu and Gul 2000) but remains underdeveloped in the legal literature and has not yet been applied in a systematic way to the central questions of international law. Some parts of the discussion that follows, therefore, will revisit notions of reputation that are familiar to some readers. Other parts, however, are new insights into how reputation can increase the rate of compliance with international law.

The simplest intuition about reputation is that a state that complies with international law will develop a good reputation and be viewed as a good partner, while a state that fails to comply with international law will have a poor reputation and be viewed as an unreliable partner. More concretely, a violation of a commitment today will affect the way a state is perceived tomorrow. A state that is known to honor its commitments will find more partners when it seeks to enter into future cooperative arrangements, will be able to extract more generous concessions in exchange for its promises, and will be able to solve more problems of cooperation than will a state that has a less favorable reputation.

If an observing state knew everything about the acting state, including the extent to which it preferred gains today over gains tomorrow (i.e., its discount rate) and the value for it of all possible interactions, the observing state would be able to calculate the acting state's payoffs and accurately predict its actions. Because these things are not observable, however, observing states form a judgment about an acting state's "reputation," which represents a measure of its willingness to comply with its international legal obligations.

To understand how concerns about reputation may affect behavior requires some understanding of why states care about reputation. One could simply assume that states prefer to have a reputation for com-

pliance with international law, in which case they would be prepared to forgo present gains in exchange for an improved reputation. The problem with this approach is that it is ad hoc and does not give us a way to consider how and when a state will value reputation more or less; or when reputation may matter more or less.

So the theory developed here assumes that states have no particular taste or preference for a good reputation, but rather are concerned with maintaining good standing within the international community only to the extent that changing one's standing or reputation affects payoffs.

To illustrate how a good reputation can be valuable to a state, think of a state's reputation as an estimate of its discount rate. That estimate is made by observing states, and so it may not align perfectly with the state's actual discount rate; furthermore, the acting state may behave strategically in an effort to influence that reputation. When a state violates a rule of international law, it may suffer a reputational loss as other states adjust their beliefs. A state with a better reputation is believed to be more patient and, therefore, more willing to comply with international law and sacrifice current gains in exchange for the ability to credibly enter into cooperative arrangements in the future. Because states cannot rely on external enforcement, reputation represents one of the few ways to make promises credible. As Schelling said, "a potent means of commitment, and sometimes the only means, is the pledge of one's reputation" (1960, p. 29).

The greater a state's reputation, the more credibly it can commit to a particular course of action, the easier it is for it to enter into cooperative arrangements, the more it can extract from other states as part of a bargain, and the more likely it is that it can find other states with which to cooperate (Lahno 1995).

The intuition can be seen through the lens of basic bargaining theory. When two or more parties enter into a contract, they do so with an eye toward maximizing their joint payoffs, taking transaction costs into account. The ability to use contract is valuable to the parties, and in the domestic context relies on the enforcement power of courts. The ability to enter into agreements is similarly valuable to states, but international agreements must make do with weaker commitment devices. In the absence of coercive enforcement, they must rely on reputation as a disciplining device to encourage compliance. The stronger a state's reputation, the more easily it can make credible commitments and resolve cooperation problems.

If improving one's reputation can yield value in the form of higher payoffs, then states have an incentive to develop and maintain a good reputation. They can do so by complying with existing obligations even when, absent a concern for reputation, it might be in their interest to violate them. One can think of decisions to comply in these situations as costly signals that serve to enhance a state's reputation for compliance. The signal is effective because it distinguishes states that are likely to comply in the future from those less likely to do so.

To explore the features of reputation further, it is helpful to work with a specific example, and the previously discussed ABM case can be revisited for this purpose. I have intentionally chosen an example involving only two states so as to keep the discussion simple. The theory and conclusions reached, however, generalize to any number of states without difficulty.

The agreement can be modeled as a two-stage game.[8] In the first stage, states negotiate over the content of the law and their level of commitment. The negotiation might involve a dickering over terms, a take-it-or-leave-it offer from one party to the other, a decision to join a preexisting agreement whose terms are already set, or even a coercive negotiation in which one party has little choice but to consent to the proposed agreement. The way the negotiation takes place will have important implications for the agreement, including the terms that are included, the use of dispute resolution, the choice between hard and soft law, and more. These issues are put to one side for the moment. It is enough here to note that the parties come together and attempt to reach agreement. If no agreement is reached, the legal obligations of the parties are unchanged; for the purposes of this example (and for simplicity), I will refer to this outcome as one in which the states are unconstrained by international law. In a more realistic discussion, one could take into account the fact that states remain bound by any other rules of international law that may be in place. This simplifying assumption, however, has no impact on the analysis or the results.

In the second stage of the game, compliance decisions are made. Each party to the agreement decides whether or not it will carry out its obligation under the agreement. In this stage, there are three possible outcomes. First, the interests of the parties may lead them to comply without regard for the agreement. In an arms control context, for example, a state may choose to limit its acquisition of weapons sufficiently to be in compliance, and this decision may have nothing to do with the

agreement. There is compliance in this situation, but it cannot be attributed to the agreement. Another possibility is that the parties face payoffs that will cause them to violate the agreement. If the payoffs from defection in the ABM context are large enough relative to the payoffs from compliance, the parties violate the agreement. Finally, it may be that the agreement itself causes the parties to change their behavior and come into compliance. The agreement itself might have caused the United States and the Soviet Union to limit their acquisition of weapons. It is this last possibility that is of greatest interest. If international law matters, it must be able to generate at least some cooperative behavior that would otherwise be absent.

In the ABM example, recall that the (stylized) one period prisoner's dilemma payoffs are represented as shown in table 2.

To understand the role of reputation, it is necessary to compare present and future gains, and so we must account for the fact that states prefer payoffs today over payoffs tomorrow. Let r represent the discount rate of states, meaning that they are indifferent between a payoff of 1 today and $1 + r$ tomorrow. We assume that each state knows only its own discount rate and so must estimate the discount rate of other states on the basis of their observed behavior. For simplicity the variable r will be used to refer to the discount rate of both states, which implies an assumption that both have the same discount rate. Strictly speaking, each state should have its own discount rate which could be achieved by adding an appropriate subscript. I omit this bit of notation in the interests of simplicity. It has no impact on any of the results or the analysis.

The estimate of another state's discount rate can be thought of as its reputation. As I discuss later, reputation refers to more than just the discount rate, but that is a useful place to start.

The total value of the payoffs to states if they both defect in every period, then, is $80 + 80/(1 + r) + 80/(1 + r)^2 + \ldots = 80(1 + r)/r$.[9] This is what each state receives in the absence of cooperation.

Table 2. The ABM Prisoner's Dilemma

		Soviet Union	
		Comply	*Violate*
United States	Comply	(100, 100)	(−50, 200)
	Violate	(200, −50)	(80, 80)

For an international agreement to influence outcomes, it must be that the violation of that agreement generates some form of cost. If reputational sanctions can increase the costs of noncompliance, cooperation is possible.

At the time of the ABM talks, the United States and the Soviet Union both stood to benefit from an agreement that would reduce the burden of their arms race. If both sides respected the agreement, they would each receive a payoff of 100 in every period. Using the same calculus as above, the discounted value of this outcome is $100(1+r)/r$. This is clearly a better outcome for the parties than mutual defection. The problem is that in any given period, each party is better off if it defects. Before choosing to violate the agreement, however, a state must consider the impact of defection on future payoffs. If violation generates future costs, these costs may be enough to bring about compliance.

How can a violation generate costs? In the next chapter I present a more detailed discussion of how and when reputations change, but for the moment it will simply be assumed that when a state violates a commitment, other states take note and the violating state's reputation suffers. When the state enters into a treaty, it represents to the other party that it prefers mutual cooperation to noncooperation. Its hope is that this claim will be credible to the other state. Framed in terms of the discount rate, the state asserts that its discount rate is such that it will cooperate. If that claim is credible, the other state will enter into the agreement. If the state then violates the agreement, its ability to make credible promises in the future will be reduced. This is what is meant by a loss of reputation—the state is less able to make credible claims about its willingness to comply (illustrated here through claims about the discount rate).

The loss of reputation matters because it makes future promises less credible. Potential partners will have less confidence that the state will resist opportunities to violate the agreement and capture some immediate gain. In this example, the reputational harm will undermine future arms negotiations and may make it more difficult to enter into other kinds of agreements as well. To the extent it does so, the gains the state can capture through such future agreements are reduced.

So as a state considers a violation of the ABM agreement, the costs of breach must be taken into account. By violating today, the state gets a payoff of 200 rather than the 100 it would get if it complies (assuming the other state complies). As a consequence of the violation, however, the state's reputation is harmed, and future negotiations on arms control issues (and perhaps other things) are undermined. Assume that this leads to payoffs of 80 in each future period. So the payoff from a violation today is: $200 + 80/(1+r) + 80/(1+r)^2 + \ldots = 120 + 80(1+r)/r$.[10]

All of the foregoing payoffs can be represented in a single figure. Because the parties know the consequences of their decisions and the associated payoffs, and because the game does not change over time, each state will either comply in every period (assuming the other side does the same) or defect in every period. To simplify the presentation, let $R = (1+r)/r$. Notice that R is always greater than 1, and as the discount rate increases (i.e., states care less about the future), R gets closer to 1. The resulting payoff matrix is shown in table 3; the headings refer to the state's action in the first period. In subsequent periods, there is mutual compliance if and only if that is what took place in the first period. If either side violates in the first period, there is mutual violation in all future periods. The payoffs reflect the discounted value of all present and future payoffs.

It is clear from this table that for certain values of *R*, both states have an incentive to comply with the agreement. In particular, as long as $100R > 120 + 80R$, there is a stable equilibrium in which both states comply. Put another way, as long as the states care enough about the future, mutual compliance can be achieved in this game. In this particular example, the discount rate, r, required to generate cooperation is 0.2.[11] In other words, as long as the parties prefer a payoff of 1.2 tomorrow over a payoff of 1 today, cooperation can be sustained.

This illustration shows that rational states can use international agreements to resolve a prisoner's dilemma. Cooperation is possible because a failure to honor the agreement affects the expectations and behavior of other states. When a state fails to comply in one period, other states observe this and draw negative inferences about the likelihood of

Table 3. The Multi-period ABM Prisoner's Dilemma

		Soviet Union	
		Comply	*Violate*
United States	Comply	(100R, 100R)	(80R−130, 120 + 80R)
	Violate	(120 + 80R, 80R−130)	(80R, 80R)

future compliance. That is, the reputation of the violating state is diminished. This model of compliance has the appealing feature that violations are "punished" through these sanctions but the sanctions themselves are not costly to the sanctioning states.

So when states enter into international agreements, they are in effect pledging their reputation as a form of bond. If they violate the agreement, they give up some of this reputational collateral, and this fact both increases the likelihood that they will comply and makes their promises more credible. The greater a state's reputational collateral, the more credible are its promises, and the easier it will be to achieve cooperation. A state with a great deal of reputational collateral will find partners easily and will be able to extract larger concessions from them. The result of this logic is that states will at times be prepared to forgo short-term opportunities to violate the law and extract higher payoffs in the hope of building or preserving their reputations and thereby enjoying higher payoffs later.

If states do not value their reputations, of course, no incentive for compliance is generated. Though this fact suggests that there are some limits on reputation's ability to generate cooperation (as I discuss in the next chapter), there seems to be nearly universal agreement that states are, indeed, concerned with their reputations. Even critics of reputational theories in political science (Mercer 1996, pp. 19–25) and international law skeptics (Goldsmith and Posner 2005, p. 103) concede as much.

Notice that the potential for cooperation is sensitive to the underlying payoffs in the game. In the foregoing example, some payoffs require unrealistically low discount rates in order to sustain cooperation (indeed, for some payoffs no positive discount rate will sustain cooperation) while for others cooperation will emerge with a discount rate that plausibly captures the trade-off states make between the present and the future.

So reputational sanctions can generate an incentive to comply with international legal obligations and can lead to changes in behavior, even in the absence of formal enforcement mechanisms. Like any other incentive, this one operates at the margin, and will at times be too small to affect state behavior. When a state is deciding whether to comply, it will take into account a variety of cost and benefits unrelated to law—domestic interests, political relations with other states, and so on—but

it will also consider the legal implications of a violation. The likelihood and magnitude of reputational sanctions will vary depending on the context, so states must assess them on a case-by-case basis. Because international law increases the costs of a violation, it puts a thumb on the scale in favor of compliance or, as is sometimes said, generates "compliance pull."

Even when breach generates a reputational sanction, it may be in the interest of the country to violate its obligations. In the earlier illustration, for example, if the discount rate of American or Soviet leaders is too high, cooperation will fail. Social scientists have referred to expectations of future cooperation as the "shadow of the future."[12] Sanctions that take the form of reciprocity or retaliation can lengthen the "shadow of the future," in the sense that they increase the future cost of today's violation. Reputational sanctions have the same effect; however, in contrast to direct sanctions, which are contingent on explicit reactions by other states, reputational sanctions (or reputational benefits) are the byproduct of the information produced by a state's compliance decision.

Before proceeding, a note about the theory of repeated games is in order. Once we acknowledge that states interact repeatedly over time and with no known final period, it is possible to demonstrate that this repetition can, by itself, sometimes be enough to allow states to overcome a prisoner's dilemma. This form of cooperation can take place without any international law or any exchange of promises, and can take place even if both states have complete information. The model of reputation developed here is something different. It seeks to address the question of whether international law can affect behavior. Can, for example, the signing of a treaty affect behavior and outcomes even when the treaty itself is just a piece of paper? For international law to matter in this way, it must affect the payoffs of the states. The reputational theory explains how this can be. International law implicates a state's reputation and, therefore, its ability to make credible commitments in the future. This theory only applies if there is some informational asymmetry among states. By entering into an international law agreement, states are signaling something about their discount rate, the nonreputational costs they face, or the relevant reputational costs.

So although some cooperation can be achieved simply because a game is repeated indefinitely, this book is focused on how cooperation can be further enhanced through pledges of reputation.

Reciprocity

Though there is a close relationship between reputation and reciprocity (discussed later) this book often talks of the two concepts separately. This is done for three reasons. First, reciprocity plays an important role in international law that is distinct from reputation and sometimes operates more reliably than reputational sanctions. Second, reciprocity is treated as different from reputation in many discussions. By separating it from the general discussion of reputation, I hope to engage that literature more directly. Third, reciprocity is often considered a separate issue in traditional international law discussions. The Vienna Convention on the Law of Treaties, for example, provides that "[a] material breach of a bilateral treaty by one of the parties entitles the other to invoke the breach as a ground for terminating the treaty or suspending its operation in whole or in part."[13]

Reciprocity can serve as a powerful compliance-enhancing tool in the right circumstances. In the bilateral context, where reciprocity is most effective, it is often sufficient to generate cooperation in a prisoner's dilemma. The conditions for success are the same as in any prisoner's dilemma. If the parties make compliance decisions repeatedly over time, if the gains from a single defection are small relative to the gains from cooperation, and if acts of violation are visible, then the threat to terminate one's own compliance can be enough to induce compliance by the other party. The intuition is fairly straightforward. If the parties both prefer mutual cooperation to mutual defection, and if defection today will undermine future cooperation, then a state is really choosing between ongoing cooperation and a one-time opportunistic gain followed by noncooperation.

To illustrate, consider the Boundary Waters Treaty entered into by the United States and Canada in 1909. The treaty established that the boundary waters of Canada and the United States would remain open to commercial ships of both states, regulated the diversion of these waters, and established the International Joint Commission, which was granted jurisdiction over certain matters pertaining to the obstruction and diversion of water and charged with the preparation of reports on matters of concern to the governments.[14] The underlying problem the treaty sought to address had all the features of a prisoner's dilemma. Each country had an incentive to divert water on its own side without regard for the other's interests, but if they did so, both would be worse off.

The Boundary Waters Treaty succeeded in generating cooperative behavior by both parties, and continues to do so today. In fact, compliance has come about despite the absence of mandatory and binding dispute resolution.[15] What explains the success of the treaty is the fact that each side can credibly threaten its own withdrawal if the other side fails to comply. So Canada chooses to comply with the terms of the treaty because its failure to comply would in all likelihood cause the United States to halt its own compliance. The United States has analogous incentives. Mutual compliance is enforced by a credible threat of nonperformance.

This sort of cooperation is familiar and intuitively pleasing, but one might nevertheless ask why mutual compliance is a stable outcome. Why can't Canada violate the treaty opportunistically and then reaffirm its commitment to the treaty and once again promise to comply? It is true that the United States would like to threaten withdrawal as a mechanism to ensure Canadian compliance, but there is no way for the United States to commit itself to such a course of action. Even though violation by one party absolves the other of any legal obligations, the parties are free to renegotiate the treaty after the Canadian breach. So after violating the treaty, Canada could simply state that it plans to comply in the future. Our intuition is that such a promise will often fail to persuade, but why? Canada's problem is that the credibility of its promise to comply with the treaty is reduced by the earlier violation. If Canada violates the treaty once, it will be more difficult to persuade the United States that promises of future compliance should be taken seriously. And even if the United States does accept Canada's assurances, future Canadian violations are more likely to lead to termination by the United States. The point here is that the violation undermines a state's reputation with respect to the particular agreement and treaty partner at issue. Though the United States was willing to rely on Canadian promises at the time of the treaty's signing and ratification, it would be less willing to do so after Canadian violations have taken place.

It is especially easy to appreciate the way reputation affects behavior in this context. The parties entered into the treaty because each believed it to be in its interest. In the face of a violation by Canada, that calculus may change. There are two reasons why the United States might exit the agreement. The first is at issue in the prior discussion—expectations with regard to Canadian compliance are changed as a result of the violation. The United States can infer from the violation that Canada's

incentive to breach is sufficiently strong to compel it to ignore its commitment and, therefore, Canada will likely ignore the commitment again in the future. The breach prompts the United States to update its beliefs about Canada's reputation and its (reputational and nonreputational) payoffs. With this updated information, the United States may calculate that the expected payoff from termination is greater than the expected payoff from compliance.

The second possibility is that the United States may terminate the treaty even if it expects Canada to comply in the future. It may do so to establish its own reputation for being intolerant of violations. This decision would be costly to the United States, because it would be turning its back on a cooperative agreement that made it better off (as evidenced by its original agreement to commit) but might nevertheless be worthwhile if it led to more Canadian compliance in other areas. That is, by exiting the treaty, the United States signals to Canada that it will not put up with violations of this sort and, having received this signal, Canada is deterred from acting opportunistically in other situations. I discuss retaliatory actions such as this one in detail in the next section.

In either case, the threat of termination will be sufficient to discourage Canada from violating the agreement as long as the long-term gains from cooperation outweigh the gains from a one-time violation. This appears to be the case with respect to the Boundary Waters Treaty, which remains in force almost 100 years after its signing, with both sides continuing to operate within its confines.

Though a withdrawal of reciprocal compliance may be sufficient to generate compliance with the Boundary Waters Treaty, this will not always be the case. One could imagine that at some point one side—say Canada—will decide that it could achieve higher payoffs by violating the treaty. In other words, although the treaty may have been beneficial for Canada at the time it was entered into, that may change at some later date, and at that point the threat of a reciprocal withdrawal of compliance will be insufficient to prevent a breach by Canada.

Assuming that this breach by Canada is undesirable, in the sense that the cost of breach to the United States exceeds the gains to Canada, additional sanctions would reduce the likelihood of an inefficient breach. The presence of reputational sanctions can serve this function. Even if, as seems likely with respect to most international law agreements, reputational sanctions are insufficient to generate an optimal incentive structure for states, they will often move the system in that direction.

Reciprocity will also fail to induce compliance when a threat to withdraw one's own compliance either lacks credibility or is of no consequence to a potential violator. Virtually every important human rights agreement, for example, must rely on an enforcement mechanism other than reciprocity. For concreteness, think of the ICCPR, one of the most important multilateral human rights treaties. Among its requirements is a ban on the ex post facto application of criminal laws (art. 15). Suppose one member of the treaty, say Russia, is tempted to violate this commitment. Whatever other forces might be at work to encourage compliance, reciprocity is almost certainly irrelevant. For many member states, it is inconceivable that they would respond to a violation by terminating their own compliance with the obligation not to apply criminal laws ex post facto. New Zealand, for instance, complies with that obligation for reasons having nothing to do with Russia's compliance. There is no circumstance in which a violation by Russia would constitute a reason for New Zealand to change its domestic policies on the question. Furthermore, even if reciprocity were imaginable, it is unlikely that Russia would care about the loss of this human right in other states. So even if New Zealand could credibly threaten to terminate its own compliance, this would have no impact on Russia's payoffs.

Efforts to ban nuclear tests represent an example in which reciprocity works in some instances and not others. As between the United States and the Soviet Union a commitment not to conduct nuclear tests benefited from the threat of reciprocal noncompliance. Compliance with the Limited Test Ban Treaty, for example, was supported by the credible threat that if either the United States or the Soviet Union violated the treaty, the other would do so as well.[16] Notice the contrast between that situation and the role of deterrence with respect to Iran's nuclear ambitions. Iran is a member of the NPT and under that treaty has a legal obligation as a non-nuclear-weapon state (NNWS) (as defined in the treaty) to refrain from acquiring or manufacturing nuclear devices.[17] Whatever other forces might cause Iran to comply with this provision, reciprocity does not play an important role. A violation by Iran is unlikely to trigger reciprocal noncompliance by other parties, and even if it did, it is not clear that this would represent a significant cost to Iran.

Retaliation

Returning to the example of the Boundary Waters Treaty, imagine that despite Canada's breach, the United States still believes that Canadian promises of future compliance are credible. That is, the American view of Canada and of the expected payoffs from the treaty is the same as it was when the treaty was first signed. Canada's violation gives the United States the choice of terminating the treaty or allowing it to continue in force—essentially the same position the United States was in when it decided to sign and ratify the treaty. Despite these similarities, the United States may choose to retaliate by terminating the treaty. Why would it do so?

More generally, why would any state use "retaliatory" sanctions, defined as explicit and costly punishments imposed by an aggrieved party against a violator? These sanctions, because they impose a cost on the retaliating state (as well as the violating state), differ from reputational sanctions or reciprocity, and raise the question of why a rational state would bear the cost of the sanction.

Explaining this behavior requires us to consider once again the reputational consequences of state actions. In this case, however, we must consider a slightly different reputational concern. By hypothesis, the United States is not reacting to a change in the reputation of the breaching party (Canada). Rather, the United States is acting to build or preserve a different sort of reputation—one for dealing harshly with those that violate their legal obligations to the United States.

To see how retaliatory sanctions might work, remember that a rational state would not take action simply out of spite or anger, so there must be some other reason for it to bear the cost of punishing a violator. This implies that retaliation will only take place if it generates some payoff to the retaliating state. A retaliating state is communicating to the violating state and, potentially, to other states, that it will react when its legal rights are compromised.[18] If successful, the act of retaliating will enhance the retaliating state's reputation as one that punishes a violator. The impact of such a reputation, of course, is to increase the expected cost of violating an agreement with that state. By retaliating, then, the state hopes to generate its own reputational capital that will induce its partners to be more compliant with their legal obligations. In effect, a reputation as a state that retaliates against violators creates an additional

enforcement tool. For example, in 1993 President Clinton, speaking about possible actions to prevent further Serb attacks in Bosnia, explained the American interest as "an interest in standing up against the principle of ethnic cleansing. . . . If you look at the turmoil all through the Balkans, if you look at the other places where this could play itself out in other parts of the world, this is not just about Bosnia."[19] In other words, the United States could discourage future atrocities by taking action now because it would demonstrate a willingness to punish certain violations of humanitarian law.

Retaliation may also be used in response to an ongoing violation, in which case the retaliation is intended to persuade the violating state to come into compliance. Notice first that the focus here is on costly actions by the retaliating state. Actions that are not costly and that are taken in response to an ongoing violation are forms of reciprocal noncompliance rather than retaliation. Because the retaliatory action is costly, the retaliating state takes it only to convince the violating state to change its behavior or to establish a reputation for punishing violators, as discussed in the previous paragraph. But the act of actually retaliating is, itself, effective only if it convinces the violating party that further sanctions will be applied if the violation continues. The only reason to come into compliance is to avoid *future* retaliatory sanctions. So when a sanction is imposed, it succeeds only if it evidences a willingness to impose future costly sanctions.

While the reputation I focus on throughout this book is one of compliance with one's own international legal obligations, the reputation that is at work when retaliatory sanctions are applied is one for punishing those who violate their international legal obligations. The former is useful because it serves to make a state's own promises credible, while the latter is useful because it makes the threat of sanctions credible and therefore encourages other states to honor their commitments.

Retaliatory sanctions may take any number of forms. These include taking some economic action, such as refusing to permit imports from the offending state or curtailing one's own exports to that state; reducing cooperation in some other area; terminating a treaty (as in the hypothetical Boundary Waters Treaty example); and in the most extreme cases, using military force.

If retaliation can encourage compliance, one might ask why states can't simply threaten retaliation up to the point at which there is an

optimal incentive to comply. This question gets to the most important drawback of retaliatory sanctions: they are not renegotiation proof (Abreu, Pearce, and Stacchetti 1986; Downs and Rocke 1995). To be effective, the threat of a retaliatory sanction must be credible, and that credibility depends in part on the threatening state's reputation for punishing violators. Because retaliation is costly to the retaliating state, the incentive to impose sanctions at an optimal level (or to develop a reputation for doing so) is weak, often making the threat of sanctions noncredible. The close relationship between retaliation and my discussion of reputation in this book should now be clear. A state that imposes a sanction does so to build or protect a reputation for sanctioning those that fail to honor their obligations or possibly to end an ongoing violation. The state accepts a cost today in the hope of getting a larger benefit in the future. The forces at play are analogous to those affecting a reputation for compliance. For a state to impose a retaliatory sanction, for example, it must have a sufficiently low discount rate and the cost of retaliation must be sufficiently small relative to the benefits of encouraging more compliance from other states.

Another point worth noting is that sanctions can at times serve both a signaling and a deterrence function. When a violation has taken place and is ongoing, sanctions can signal to the violating state (and other states) that the sanctioning state will punish violations, but it can also encourage the violating state to bring its actions into compliance. Once compliance is reestablished, the sanction is normally removed. Knowing this, the violating state has some incentive to come back into compliance. This is the way the dispute resolution system at the WTO is intended to work. When a state refuses to comply with a ruling of the dispute resolution organs of the WTO, the complaining state is given the authority to impose sanctions. The stated purpose of these sanctions is to bring the violating state into compliance. For example, in the *Brazil–Aircraft* case, Brazil was found to be violating the subsidies rules of the WTO, and Canada was given the authority to impose sanctions equal to the full amount of the subsidy rather than the amount that affected Canada or the dollar impact of the subsidy on Canada.[20] In their ruling, the arbitrators said "a countermeasure is 'appropriate' inter alia if it effectively induces compliance."[21] This sanction provides Brazil with an incentive to come into compliance and signals that future violations will be met with further sanctions.

International Tribunals

The Three Rs of Compliance help us to understand how the repeated nature of international legal interactions can generate compliance and cooperation without a formal enforcement mechanism. The international legal system, however, does have institutions that, at least at first glance, seek to enforce international commitments. In particular, tribunals exist that look something like domestic courts and so might be thought to have some enforcement power. There are also international rules that purport to impose penalties on those that violate international law.[22] Is it possible that these institutions play a role similar to that of domestic courts? Is it possible that, at least in some contexts, international tribunals or other international rules can provide an enforcement mechanism for international law that operates separately from the Three Rs of Compliance?

The short answer to these questions is almost certainly no. International courts and international rules calling for compliance or enforcement, though they likely play a role, cannot by themselves be said to offer an explanation of how international law promotes cooperation.

The critical difference between domestic and international courts is that the former are backed by the state and a system of coercive enforcement. A contract between two parties, for example, represents a credible exchange of promises because the state stands ready to enforce the agreement. A party who refuses to comply with her obligations can be forced to do so or to pay damages. The ability to call on coercive enforcement changes the game from a one-shot prisoner's dilemma into a two-period game in which the promises made in the first period are enforced in the second. This enforcement leads to a sanction being imposed on the violating party that alters the payoffs of the parties (relative to the simple one-shot game) and, if the sanction is large enough, generates compliance. International tribunals lack this ability to summon coercive enforcement, and so they ultimately must rely on the same Three Rs of Compliance I have already discussed. This alters the institutions as well as the way we should analyze them.

International adjudication or arbitration can take place before standing international arbitral bodies or in an ad hoc fashion. In terms

of international institutions there are several international bodies that have jurisdiction to adjudicate disputes among states or between states and private parties. The best known of these are the ICJ, the WTO and its mandatory dispute resolution system, the European Court of Human Rights, a number of human rights tribunals, the Law of the Sea Tribunal, and the arbitration bodies authorized to adjudicate disputes under bilateral investment treaties.[23]

It is worth noting that most international agreements exist without any form of dispute resolution. Agreements might be entirely silent on the question of dispute resolution or might include the singularly unhelpful command that the parties work together to resolve the dispute. For example, the Chemical Weapons Convention includes the following dispute resolution provision: "When a dispute arises between two or more State Parties . . . relating to the interpretation or application of this Convention, the parties concerned shall consult together with a view to the expeditious settlement of the dispute by negotiation or by other peaceful means of the parties' choice."[24]

Furthermore, when tribunals exist, they typically do not have busy dockets. The ICJ has handled 110 contentious cases over its more than 60-year history, the WTO and its predecessor, the GATT has received a total of approximately 650 cases over 60 years, and the International Tribunal for the Law of the Seas (ITLOS) has handled 13 cases in its 12-year history. The precise number of disputes brought under bilateral investment treaties is unknown (because such disputes are not always public), but as of 2004, 160 claims were known to have been filed (UNCTAD 2004). The most notable exception is the European Court of Human Rights, which has seen over 8,000 admissible complaints in its almost 50-year history. The large number of cases before this court is attributable to the fact that private parties are able to bring complaints. Indeed, Scott and Stephan (2006) consider the ability of private parties to initiate proceedings to be one of the hallmarks of an emerging system of "formal enforcement" of international law. Where access to a tribunal is limited to states, the number of cases is consistently modest. Nevertheless, dispute resolution is important to the international system. It is an option available whenever states draft an agreement, and it serves a prominent role in certain issue areas, most notably trade and human rights.

Where they exist, international tribunals often have the look and feel of domestic courts. They often call themselves courts, the adjudicators

are often referred to as judges, they feature an adversarial system, they rely on legal arguments, many of them publish reasoned opinions that resemble domestic court rulings, and they are charged with issuing legally binding rulings intended to resolve disputes. These similarities are motivated by a desire to have international courts serve a function similar to that of domestic courts. And perhaps because of the superficial similarities, analyses of international tribunals make frequent analogy to domestic court systems. The domestic system works well, it is said, because courts stand ready to enforce the laws. Enforcement is a central problem for international law, and the establishment of international courts, the argument goes, will make international law more effective. Indeed, the establishment of international courts and tribunals is often said to move the international system from anarchy to order and from politics to law. For example, when the WTO was established and its dispute resolution system put in place, prominent commentators claimed that under the new system, "right perseveres over might" (Lacarte-Muro and Gappah 2000, p. 401).

Such claims cannot be evaluated until we have some understanding of what international courts do and how they are relevant to the international legal system. To understand what courts add to the international system, begin with the fact that when a state loses before a tribunal there is no formal legal structure in place to enforce the ruling.[25] The assets of the noncompliant state will not be seized, nobody will be arrested, and under existing rules the state will not even lose the ability to file its own complaints on other issues before the same tribunal. If these tribunals are effective, then, it must be for some reason other than the system of coercive enforcement that goes with a domestic court's decision.

But if international courts are unable to enforce their decisions, why do they exist at all? What role do they serve? How do they make international law more effective, if in fact they do so at all? Whatever impact international courts have must be the product of changes to the payoffs generated by the ruling itself rather than by associated enforcement mechanisms. The ruling itself, however, is just that—a ruling. Its only possible role, then, is as an information mechanism. Recognizing that international courts serve almost exclusively to provide information changes the way one views and evaluates them.

There are two kinds of informational dissemination that might make an international court effective. First, it may assist the states to

come to a common understanding regarding relevant facts or law, and thereby assist in settlement. It may achieve this through a conventional adversarial process, or it may do so through something more akin to mediation. In either case, once the parties have a shared understanding of relevant facts and law, they may be able to reach a settlement that was previously unavailable. Standing courts such as the ICJ are particularly well positioned to clarify the law for future cases both because the judges remain in place over time and because in each case the tribunal is recognized as the ultimate authority on the interpretation of the relevant laws. Though the decisions of these tribunals lack formal precedential authority, judges routinely look to prior cases for guidance and treat them as settled law. The decisions of these international courts, then, reduce uncertainty surrounding the relevant legal rules.

The second form of informational dissemination an effective court might engage in is closer to what domestic courts are normally thought to do. Here the desired impact is not so much to provide information that helps the parties negotiate a resolution as to sanction a party that has violated the law. To do so, the court must be effective in distinguishing states that have violated the law from those that have not. Doing so allows the parties and other states to form a more accurate assessment of challenged behavior and, therefore, to update beliefs about a state and its willingness to comply with international legal commitments, to retaliate, or to reciprocally suspend compliance with an agreement. This is the function, for example, of dispute resolution at the WTO. To be sure, the actions of WTO panels and the WTO's Appellate Body often promote settlement, but they also rule on the question of whether the defendant has violated WTO law. If a state does not bring itself into compliance (or satisfy the complainant in some other way) sanctions are authorized. Wrongdoers, then, face both reputational and retaliatory consequences when they lose a case. To the extent that the panels accurately identify violations, they are able to provide a more effective deterrent to wrongdoing. To do this, it is critical that the panels be perceived as disinterested and beyond the control of the parties. To that end, the WTO appoints panelists through a process that does not allow the parties to control the decisions and actions of the panels or the Appellate Body.

The two aforementioned functions are not, of course, mutually exclusive. Providing accurate information about the facts and the law

can serve both to promote settlement and to increase the cost of violation. In some instances, however, one function will dominate the other. If the information provided is confidential, for example, this will tend to serve the interests of settlement rather than sanction. The same will be true if the parties control the adjudicators to a greater extent. In this case, the tribunal is not disinterested and is, therefore, less likely to arrive at reliable conclusions regarding the facts or the law. In the course of the proceedings, however, the parties may overcome critical informational asymmetries.

Needless to say, these different functions will lead to different observed outcomes. If, for example, the parties exercise considerable control over the tribunal, a ruling will emerge only if both parties accept it—or, in less extreme cases of party control, a ruling is more likely to have been accepted by the parties if the tribunal is less independent. When the parties have a greater ability to frustrate the issuance of a ruling they dislike, it follows that a ruling is more likely to be complied with. The ruling, in effect, resembles a settlement agreement, in the sense that the parties have significant control over the terms. Rational parties will resist a ruling unless they expect it to be complied with. It is to be expected, then, that the rate of compliance with the rulings of dependent tribunals will be high. This simply reflects the fact that both parties have signaled that they prefer the settlement to the alternative of a continued dispute.

When a tribunal is more independent, on the other hand, the final decision does not require the consent of the parties. Given that there is no coercive enforcement scheme in the background, one would expect a lower level of compliance. The losing party retains the option of ignoring a decision and living with whatever consequences come with that decision. This is, for example, what the EC has done in the *EC–Hormones* case at the WTO. Rather than abide by the decision of the WTO's Appellate Body, the EC has continued its illegal activity and lives with the WTO-approved sanctions imposed by the United States and Canada as a result.

Because international courts can serve one or both of the aforementioned functions, there is no coherent sense in which compliance rates can be used to evaluate the effectiveness of the dispute resolution process. When courts are dependent, they serve primarily to promote settlement, and effectiveness should be measured by the court's ability

to encourage the parties to end their dispute. Where a court serves to impose a sanction, its effectiveness depends on the extent to which it punishes violations or, perhaps more important, deters wrongdoing and causes states to terminate violative practices. Each of these roles has value, and each is to be preferred in some contexts and not others. It is not possible, therefore, to make sweeping claims about whether courts that tend to perform one of these functions rather than the other are in some sense "better" than those performing the other (Helfer and Slaughter 2005; Posner and Yoo 2005).

Because their primary role is informational, international tribunals can be effective even if (as is almost always the case) they are not accompanied by any sanction or enforcement authority. Dispute resolution at the ICJ, for example, provides a mechanism through which information is provided on the state of the law and the actions taken by the defendant. Though there is no enforcement, the ICJ adds value by clarifying whether a state has violated the rules of international law. This action provides information to the parties to the dispute, but it also provides information to third parties, allowing them to remain informed about the relevant behavior even when they are not directly involved (Maggi 1999; Milgrom, North, and Weingast 1990). This increases the reputational consequences of violating the underlying international legal rule, even though the tribunal does nothing more than state the actions taken by the defendant and declare whether such actions violate international law. By reducing uncertainty about legal rules and state conduct, the tribunal encourages compliance. The use of dispute resolution, then, can improve the working of an agreement, even if there is no enforcement of the judgment.[26]

It should be added that international tribunals may play an additional role in the international legal system. Rather than simply existing for the sake of resolving disputes, they may serve to establish or clarify the substantive rules of international law (Danner 2006). If so, this may be a reason, quite apart from the compliance effect of the tribunals, for their existence. Having these institutions operate in a sort of quasi–common law role may be problematic from a legitimacy and democracy perspective, but it also has the advantage that it allows gap-filling and even (to some extent) changes in legal rules. This feature of tribunals may also represent a reason (in addition to those I discuss in chapter 4) that states are at times reluctant to provide for formal dispute resolution in their agreements.

State Responsibility

International law also includes some rules that at least arguably require states to provide compensation to a state that has been wronged.[27] For example, article 31 of the Draft Articles of State Responsibility states that "[t]he responsible State is under an obligation to make full reparation for the injury caused by the internationally wrongful act."[28] This is said to be a codification of a customary international law regarding reparations and can be traced to the *Chorzow Factory* case, decided by the Permanent Court of International Justice in 1927.[29] Even if one assumes that the rules of state responsibility have an effect on states, it remains to be explained why this is so. In trying to understand why a state might comply with an international obligation, it makes no sense to turn to a rule of international law that says a failure to comply generates an obligation to make reparation. If there is nothing else to encourage compliance with the initial obligation, then the rule requiring reparations will be similarly impotent.

In game theoretic terms, any mechanism that induces cooperation must ultimately do what the court system does in the domestic context—it must alter the payoffs received by the states in such a way as to encourage compliance. In other words, there must be some way the decision to violate international law imposes costs on the state. A desire to avoid these costs is what provides the incentive to comply with international law.

Payoffs and Strategies over Time

Up to this point, the discussion has assumed that the payoffs facing the parties remain stable over time. By this I do not mean that they are the same in every period (which is not needed for the analysis) but rather that there are no external shocks that change the payoffs in future periods. This assumption of stable payoffs is convenient, but it is more realistic to assume that payoffs change over time, and often do so in ways that are difficult to anticipate. Returning once again to the ABM Treaty example, the decline of the Soviet Union in the 1980s, and its dissolution and the subsequent weakness of Russia during the 1990s, increased the incentive for the United States to violate the treaty, and ultimately led to its withdrawal from the treaty. Much of the incentive

for the United States to comply with the ABM Treaty was driven by the desire to avoid an arms race. Once it became impossible for the Soviet Union/Russia to pay for an arms race, the United States stood to gain from the building of an antiballistic missile system. Russia would simply be unable to counter U.S. technological advances. Meanwhile, U.S. officials felt that U.S. technology had improved sufficiently to make viable an ABM system designed to deter threats from rogue states that had acquired or were seeking to acquire nuclear weapons. This combination of changed circumstances was not something the parties could have anticipated with any confidence, and so their efforts at cooperation were not specifically tailored to account for this confluence of events.

If states know that shocks may occur, they can at most calculate the expected future payoffs. These, however, may differ substantially from the actual payoffs facing the states in a future period when they make their compliance decisions. With only estimates of future payoffs, states may not know whether they will be playing a coordination game, a prisoner's dilemma, or some other game.

Relaxing the assumption of stable, known payoffs yields at least two insights. The first is that a country's decision to comply with a rule at one moment in time does not necessarily imply that it will continue to do so at some other time. For example, a state that is enjoying good economic times may decide to comply with an environmental agreement because it is willing to give up some economic benefits in exchange for improved environmental practices. If that same state finds itself in a recession, however, it may weigh the economic costs of the agreement more heavily, and may decide to violate its commitment. In both the good and the bad times, that state has behaved rationally to maximize its payoffs. The outcome has changed because the nonreputational payoffs have changed.

This illustration suggests a definition for reputational and nonreputational payoffs that I will use throughout the book. Reputational payoffs are generated by changes to a state's reputation. The reputational changes help other states to predict what the state will do in the future. Nonreputational payoffs are those payoffs that are the direct result of a state's decision to act in one way or another. They are independent of how the behavior affects the state's reputation.

The second insight is that states engaged in what appears to be one sort of game may enter into agreements designed to address problems

of cooperation in some other game. States concerned about a future shock may structure their cooperation to deal with both the game as it appears on the basis of today's payoffs and the game they may find themselves playing after a shock. This offers a possible (though only partial) explanation for why states sometimes use international law to resolve games that do not appear to require any form of enforcement mechanism (e.g., games of common interest, coordination games, battle-of-the-sexes games). If future payoffs are known at the time an agreement is reached, cooperative games of this sort can normally be resolved at lower cost by using less formal ways to select a focal point. For example, a unilateral declaration by one of the states will often be sufficient to generate a focal point and establish a stable equilibrium.

But if the game in question, though it looks like a coordination game, has some probability of becoming a prisoner's dilemma or some other game in which cooperation is more difficult, the states have an incentive to protect against that outcome. By way of example, the Antarctic Treaty bans the establishment of military bases and the testing of weapons on Antarctica, suspends territorial claims, sets up an inspection system, and provides for periodic meetings of the parties. When it was signed in 1959 (and entered into force in 1961), the prohibitions included in the treaty had little practical effect, because states were generally already in compliance. There were, at the time, no military operations in Antarctica, territorial claims were not immediately threatening to escalate into military conflict, and the territory had little more than speculative strategic value. In this sense, the treaty could be described as an effort to resolve a coordination game. The states involved wanted to preserve the territory for scientific purposes and keep it free of military activity and weapons testing, but it seems likely that this was precisely the equilibrium that had already been reached.

Looking forward from the time of the signing, however, it is plausible that the parties had concerns about how the importance of Antarctica and therefore the payoffs to the parties might change. If the interests of the parties changed, whether for economic (e.g., the discovery of oil or mineral reserves), strategic, or other reasons, the game might become a prisoner's dilemma. By establishing a treaty rather than a more informal set of norms, the parties solidified the cooperative

regime. As it turned out, environmental issues have become important in Antarctica, and these have the character of a prisoner's dilemma. Although the original text of the treaty contains no specific references to environmental protection, it quickly became clear that environmental issues were among the most important.[30] The first of a series of additional agreements and protocols on the environmental protection of Antarctica was signed in 1964, only three years after the treaty went into effect. Furthermore, in 1988 the Convention for the Regulation of Antarctic Mineral Resource Activities (CRAMRA) was adopted. Though this agreement contained stringent environmental protections, it nevertheless permitted mining in Antarctica, indicating that states' interest in mining had increased to the point where cooperation under the auspices of the original treaty was necessary to solve the nascent prisoner's dilemma. In fact, the reaction to permitting any mining in Antarctica, however limited, was so negative that CRAMRA was superseded in 1991 by the Protocol on Environmental Protection to the Antarctic Treaty (the Madrid Protocol). The evolution of these environmental agreements illustrates the wisdom of using a formal treaty to establish rules governing Antarctica: as the need for more explicit environmental regulation of Antarctica grew, the consultative bodies established by the original Antarctic Treaty provided a forum in which states could bargain to solve the emerging prisoner's dilemma.

That states may enter into international agreements to generate cooperation in games with payoffs that change over time suggests that in some situations international law is doing more work than it seems. So even if many international agreements, at the time of their formation, function primarily to resolve coordination games, international law may nevertheless be generating cooperation in games that, because of changing payoffs over time, represent more difficult cooperative problems.

Modulating the Level of Commitment

The following discussion of the level of commitment is closely related to the theory developed in Guzman (2002b) and Sartori (2002). In Sartori's model, states issue costless signals about the importance of particular issues. Despite being costless, these signals matter because when a state signals that an issue is of great importance, rivals understand that attacking the state on that issue is more likely to lead to

resistance and so are less likely to attack. Though a state could claim that every issue is important to it, doing so would be a poor strategy because this would cause its statements to lose credibility. Instead, rational states admit (at least much of the time) when an issue is of only modest importance because doing so lends credibility to its statements about issues it does consider critical. The result is that states use costless signals to send different messages reflecting their commitment.

In the international law context, states can send different signals about their intention to comply with international legal obligations by using different legal forms. They can refrain from any international agreement, enter into a "soft law" agreement (defined as an agreement that is neither a formal treaty nor customary international law), or enter into a formal treaty.[31] Within the category of soft law, further distinctions are possible. For example, the United States has entered into a number of soft law agreements addressing matters of cooperation in antitrust. These agreements demand very little of the participating states, whereas the Basel Accord, also a soft law agreement, includes quite detailed provisions about what states must do to comply and requires real changes in the practices of the states involved. By choosing one form of commitment over another, states signal their seriousness and the amount of reputational collateral they wish to pledge. A formal treaty represents the most serious form of commitment not because it is more costly to enter into than other commitments but because it is understood to be a maximal pledge of reputation (Guzman 2002b). There is more to be said about the role of soft law and treaties, and that discussion is in chapter 4.

In game theoretic parlance, the choice among different forms of agreement is cheap talk. Even if it is no more expensive to use the treaty form rather than soft law, states may do so to signal a more serious commitment. Choosing between the treaty and soft law forms is not so different from choosing to write an agreement on red paper when states wish to signal a strong commitment and green paper when they wish to signal a weaker commitment. By making a more serious commitment (e.g., a treaty) a state is able to extract more from its counterpart because the promises it makes are more credible. The temptation to make every agreement as formal and serious as possible is offset by the costs that come with such a commitment. Specifically, if the agreement is violated, the violating state suffers a reputational sanction. Where the commitment is less serious, then, the state has an incentive to make this clear by, for example, using a soft law form rather than a treaty.

Coercion and International Agreements

Throughout most of this book, the discussion assumes that international agreements are consensual. Under current international law rules, a treaty is considered void as a result of coercion only if the agreement was the result of the unlawful use or threat of force or if a country's representative was threatened (Vienna Convention on the Law of Treaties, arts. 51, 52).[32] Analytically, these rules seem to leave plenty of room for agreements that could be described as coercive, but it turns out to be difficult to craft a better rule to distinguish agreements that increase the well-being of both states from those that do not.

One way we can be sure that an agreement is Pareto improving (meaning that it makes some parties better off without making any worse off) is to demand that all parties have the option of rejecting the agreement and retaining the status quo. Strictly speaking, because individuals make these decisions on behalf of states, consent only ensures that these individuals prefer the agreement to the alternative of no agreement. If these decision-makers pursue private objectives that are inconsistent with the general welfare of their citizens, even consensual agreements need not improve welfare. Nevertheless, we expect consensual agreements to improve the lot of the parties involved with greater frequency than nonconsensual, coercive agreements. After all, citizens (at least within democracies) have at least some check, through the ballot box, on the international activities of their politicians.

There is no doubt that at least some international agreements are, indeed, coercive. An obvious example is the 1919 Treaty of Versailles, by which Germany surrendered to the Allied Powers following World War I. As is true of many other peace agreements, there is no sense in which that agreement can be described as being entered into voluntarily by the German government, nor could Germany choose the status quo rather than the proffered agreement. Agreement was achieved at gunpoint.

Plenty of other, less obvious examples of coercive agreements exist. In September 2004, for example, the United States signed the Trade and Investment Framework Agreement with Afghanistan. At the time, and indeed still today, the government of Afghanistan relied so heavily on U.S. support that the decision to sign an agreement presented by the United States can hardly be construed as a free choice. Another example from the early twentieth century is the Hay-Bunua-Varilla Treaty of

1903, which gave the United States the Panama Canal Zone and the right to build the Panama Canal. Having successfully declared independence from Colombia (at the urging of U.S. president Theodore Roosevelt), the new nation of Panama was in need of allies in the region. The powerful allure of the aid and protection of the United States against a potentially hostile neighbor, Colombia, surely exerted a coercive influence on the negotiation of the treaty that granted to the United States one of the most valuable property rights in the world.

Other agreements, however, are more difficult to classify as coercive or consensual. A conspicuous and important modern example is the agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPs), which was entered into by all WTO member states at the conclusion of the Uruguay Round of trade talks in 1994. The agreement requires that the signatories establish a domestic regime with certain intellectual property protections. Critics of the agreement point out that most developing countries had almost nothing to gain from the TRIPs agreement, and a great deal to lose. They argue that these states signed the agreement because they were threatened with a withdrawal of access to the markets of developed countries, most notably the United States.[33] Defenders of the agreement respond that it was part of a larger bargain struck by states during the Uruguay Round of trade negotiations in the late 1980s and early 1990s. The TRIPs agreement was just one part of the WTO agreement that emerged from negotiations that included concessions by developed states on agriculture, textiles, and the use of unilateral trade measures. Though it was the result of hard bargaining, it was not coercive, the argument goes, because all parties got something they valued.[34] Who is correct in this exchange? Almost certainly both sides. On the one hand, there is no question that the negotiation of the Uruguay Round was a bare-knuckles affair in which power played an important role. Then again, this is true of virtually all important international negotiations. If the presence of large differences in bargaining power is enough to make an agreement coercive, then virtually any north–south agreement (among many others) qualifies. On the other hand, at some point hard bargaining turns into an exchange of threats. The final WTO agreement was described as a "single undertaking" that required states to accept all the negotiated rules or remain entirely outside the system. Furthermore, the previous GATT regime was effectively dissolved by the new WTO, so states declining the new WTO did not have the option of retaining the old status quo.

Developing states were thus forced to choose between accepting the entire WTO package, including the TRIPs agreement, or remaining outside the newly formed WTO and therefore outside the trading system's most important institution. Because states did not have the option of retaining the pre-WTO status quo, the agreement satisfies my definition of coercion. The problem is that there may not have been any other way to reach agreement, and it is possible that every state was made better off by the WTO agreement. The complex package of concessions that emerged from the Uruguay Round negotiations was fragile. If states were permitted to commit to some provisions but not others, the entire structure threatened to break down as each state withdrew the concessions it had made. And if states could opt out of the new WTO while retaining all their rights under the GATT, there would be a temptation to free ride on the additional liberalization provided for under the WTO. By giving up (or being denied) the option of retaining the status quo, states may have improved their payoffs. The single undertaking, then, can be viewed as a necessary mechanism to preserve the integrity of the agreed-upon terms.

One is tempted, then, to rethink the definition of coercion offered at the start of this section. Perhaps agreements should sometimes be considered consensual even if the parties are unable to reject the terms and opt instead for the status quo.

This example demonstrates the key problem: to the extent that we would like to identify a set of agreements that can be labeled as coercive.and that are, for that reason, normatively problematic, there is no straightforward way to distinguish coercive agreements from consensual ones (Keohane 1984). One might seek a definition under which consensual agreements are those that (at least in expectation) lead to a Pareto improvement (i.e., make all parties better off), but there is no reliable way to determine when an agreement satisfies this definition. There is nothing unique about the international arena in this context, of course. The same problem exists in the domestic sphere where we cannot clearly identify the difference between a coercive contract and a noncoercive one. We know that "your money or your life" is a coercive agreement, but beyond that it quickly becomes difficult to establish clear distinctions.

What is the relevance of all of this to the working of international law? As mentioned, coercive agreements (defined as those that do not offer the option of retaining the status quo) are more problematic from a welfare perspective than are consensual ones. Indeed, a coercive agreement need not even lead to an improvement in total welfare, however

measured. It is entirely possible that the gains to one side will be outweighed by the losses to the other, meaning that the agreement destroys value, rather than creating it.

Despite this important difference between the welfare effects of coercive and consensual agreements, the two types of agreement can be treated similarly for compliance purposes. A coercive agreement is presumably valuable to the party doing the coercing for the same reason that a consensual agreement would be valuable. That is, the agreement must generate some compliance pull beyond what can be achieved without it. After all, why coerce a state into signing an agreement if that agreement does not affect its behavior? When a state makes a promise to take some costly action and in exchange receives some benefit or avoids some cost imposed by the other side, it has an incentive to refuse performance when the time comes. Whether the agreement is a coercive one or a consensual one, a refusal to comply may provoke reputational sanctions. It is true that the coercing state may threaten to impose some additional sanction if there is a violation (presumably this threat would resemble the one that led to the agreement in the first place), but this threat of sanction works like any other such threat—it is only effective if it is credible, and a costly sanction will only be imposed if it contributes to a state's reputation for penalizing those that violate international agreements.

There may, however, be a natural upper bound on the extent to which coercive agreements can bind a state. At least when it is clear that coercion was used, the reputational sanction for violation may be limited. A violating state able to show that it was coerced may be able to limit the reputational impact to its reputation for compliance *with coerced agreements*. The state may be able to preserve a good reputation for agreements that are truly consensual. A state's ability to cabin the reputational impact of its breach depends on the extent to which reputation is compartmentalized, a question I discuss in Chapter 3.

Multilateral Cooperation

As mentioned, both reciprocity and retaliation are more effective enforcement mechanisms for bilateral agreements than for multilateral agreements. In fact, some observers have suggested that these

sanctions are so ineffective in the multilateral context that multilateral cooperation achieved through international law is virtually nonexistent.[35] This negative conclusion about multilateral cooperation, however, both understates the potential role of reciprocity and retaliation and ignores the fact that reputational sanctions represent an alternative compliance-promoting mechanism in multilateral agreements. In this section I discuss the problem that multilateral cooperation presents for reciprocity and retaliation, explain why it is often (but not always) likely to undermine these enforcement strategies, and discuss why reputational sanctions can be effective even when many parties are involved in an agreement. The conclusion of the section is that multilateral cooperation presents some unique challenges for cooperation, but no theoretical reason exists to think that international law cannot facilitate it in a wide range of contexts.

Reciprocity and Multilateral Agreements

There are several reasons that states might choose to enter into a multilateral agreement rather than a series of bilateral ones. For example, multilateral agreements can achieve uniformity more easily than can bilateral agreements, there may be economies of scale in verification and monitoring that promote a more effective multilateral reputation mechanism, and multilateral negotiations may allow for a richer set of possible exchanges and trade-offs than do bilateral talks—potentially creating more room for agreement.

For my present purposes, however, it is sufficient to focus on just one reason: multilateral agreements allow states to overcome collective action problems that bilateral agreements cannot adequately address.

When state decisions affect all (or many) other states, as is the case with many environmental, human rights, and nuclear arms policies, among others, there are powerful reasons to address these issues multilaterally. The reasons are familiar and are illustrated by the environmental problems that gave rise to the Kyoto Protocol. Greenhouse gas emissions are harmful to the atmosphere and therefore represent a cost imposed on the entire world. The cost of reducing those emissions, however, is borne by individual states (and their private industries) that enact regulatory measures. Because the government of a state takes into account all the costs of tougher pollution standards but only a fraction of the global benefit from reduced environmental harm, states will systematically

underregulate. Multilateral cooperation allows states to act collectively and internalize more fully the costs and benefits of their policy choices.

The Kyoto Protocol provides this kind of solution. It is a multilateral agreement that calls on industrialized states to reduce their collective emissions of six greenhouse gases blamed for climate change. The treaty, which entered into force on February 16, 2005, commits ratifying states to reducing emissions or, in the event that they maintain or increase their greenhouse gas emissions, purchasing emissions credits from states that have met their targets. The problem posed by climate change, and addressed by the Kyoto Protocol, has the structure of a multilateral prisoner's dilemma. All states have an interest in the reduction of harm to the environment in general and reducing global warming in particular. Because there are many states involved, however, each state has an incentive to shirk and free ride on the efforts of other states. Absent some effective form of multilateral cooperation, there will be too little investment in environmental protection. Collective or multilateral decision-making, in the form of an international agreement, ensures that the member states more fully internalize both the costs and benefits of regulation.

The very reason that public goods problems (i.e., problems involving nonrivalrous goods, such as the global environment, for which consumption by one individual does not reduce the amount of the good available for consumption by others) are often best addressed through collective action also makes reciprocity less effective as an enforcement device. Imagine that a state fails to honor its commitments under the Kyoto Protocol. It would make little sense for all states to simply cease their own compliance in response. Doing so would undermine the purposes of the treaty and impose a cost on all states. Recognizing this, a potential violator knows that the agreement is likely to continue to exist whether or not it violates its commitment. Since every state recognizes this fact, everyone's incentive to comply is reduced. In short, reciprocity is unlikely to prove an effective tool to sustain compliance in a multilateral treaty aimed at a public goods problem.

Though this problem of reciprocity is real for many multilateral agreements, it will not always present a challenge. In agreements that address pure public goods, states are unlikely to be able to condition their own performance on reciprocity by all other states. For many other multilateral agreements, however, it is possible for a state to respond to a violation by withdrawing its own compliance only with respect to the

violating state, making reciprocity a useful and sometimes very powerful tool to encourage compliance.

An obvious example is the WTO system, in which a multilateral treaty is enforced by granting states the authority to selectively suspend their own compliance with portions of the agreement. So, for example, when the WTO's adjudicatory bodies concluded that the European Communities had violated their obligations with respect to the importation of beef containing artificial growth hormones, they granted the United States and Canada the right to suspend certain trade obligations they had toward Europe. Though not perfect, this system of reciprocal withdrawal of benefits represents a meaningful incentive to comply with the multilateral trading rules.

In general it is the case that when the good in question is excludable, reciprocity should work in the multilateral context in much the same way as it does in the bilateral context. When the good is non-excludable, the use of reciprocity to enforce multilateral regimes becomes significantly more problematic.

Retaliation and Multilateral Agreements

The threat of retaliation can often serve as an enforcement device, but like reciprocity it works less well when public goods are involved. Consider, by way of example, the ICCPR. This multilateral human rights treaty imposes obligations on all member states. Suppose Russia is tempted to violate its commitment. A credible threat to impose a retaliatory sanction—perhaps a ban on exports from other states to Russia—might affect that compliance decision, but such a threat is likely to lack credibility for two reasons. First, it suffers from the same credibility problems that exist in the bilateral context. Specifically, imposing the sanction is costly for the sanctioning party, and a rational state will only take this action if there is some benefit. As discussed in the context of bilateral agreements, one reason to impose the sanction is to acquire or protect a reputation as a state that punishes violators, but this may not be enough.

The second problem that is particular to the multilateral context is the free-rider problem. Even if the threat to sanction a violation would be an effective deterrent, when it comes time to impose the sanction, each individual state has an incentive to free ride on the actions of others. Every party to the treaty (and arguably every state in the world) benefits from compliance by Russia and therefore from the compliance

inducing effect of the threatened retaliation.[36] But only those states that impose the sanction (or threaten to do so) bear the cost of the sanction. So every state has an incentive to try to capture the benefits of compliance without bearing the costs of the retaliation, and it is likely to be difficult, as a result, to credibly threaten to sanction a violator. To be sure, multilateral sanctions are imposed from time to time. Prominent examples include the economic sanctions on the South African apartheid regime and sanctions on Iraq following the first Gulf War.

In many other cases, however, it proves impossible to overcome the collective action problem. The six-party talks associated with North Korea's development of a nuclear weapon and its withdrawal from the NPT illustrate this point. The United States, China, South Korea, Japan, and Russia were, at least for a time, unable to impose a sufficiently high cost (or opportunity cost, in terms of benefits forgone under the parties' 1994 agreement) to prevent North Korea from taking these actions.[37] Part of this problem was a failure to overcome the collective action problem and impose a unified sanctions regime. More recently the agreement reached among the relevant states promises to end North Korea's nuclear program.

The collective action problem with respect to retaliation is a real and serious one in many contexts, but in some situations states are able to overcome it. If a violation imposes costs on one (or a few) states, a threat of retaliation may be credible even if the underlying obligation is multilateral. The Vienna Convention on Diplomatic Relations, for example, imposes obligations on every state toward all other states that are party to the agreement, but when a violation occurs it normally injures only one party, and that party can retaliate. The collective action problem is overcome in this context because the retaliating state only threatens retaliation when its own interests are harmed and it is, therefore, able to capture the full benefits of a reputation for punishing such violators. The threat of retaliation in this context is as credible as it would be in a bilateral agreement. Several prominent multilateral treaties in which retaliation of this sort is possible are trade treaties (e.g., WTO; NAFTA; the Dominican Republic-Central America Free Trade Agreement, DR-CAFTA), but the same sort of sanction could exist under a variety of treaties in other areas (e.g., the Basel Accord, Geneva Convention

Relative to the Treatment of Prisoners of War, Vienna Convention on Diplomatic Relations, VCCR).

It is also worth noting that multilateral agreements in which there are only a few (or perhaps just one) leading parties, as is often the case in today's world, will face a smaller collective action problem. If the leading party captures a large share of the gains from compliance, it will have a greater incentive to impose retaliatory measures, and therefore its threat to do so will be more credible. An example of an agreement of this sort might be the NPT.[38] This treaty has 187 member states, but among these the United States plays a leading role in encouraging compliance by others. As the only global superpower, the United States' interests are affected, and its influence diminished, whenever another country anywhere in the world develops a nuclear option that negates the substantial U.S. advantage in conventional weaponry. As a result of this distribution of benefits, the United States has taken the lead in organizing, coordinating, and contributing to enforcement efforts such as the Proliferation Security Initiative (PSI), a voluntary arrangement of states aimed at halting the spread of illicit weapons technology through active and cooperative interdiction of illegal arms shipments.

The PSI, launched in 2003 following North Korea's withdrawal from the NPT, was widely seen as an attempt to halt the North Korean nuclear program. Similarly, Israel stood to lose considerably more than other states in the 1980s if Iraq developed a nuclear capability in violation of its NPT obligations. It should thus not be a surprise that Israel was willing to bear alone the costs of a preemptive strike that destroyed Iraq's Osiraq nuclear reactor in 1981.

Reputation and Multilateral Agreements

So even if one considers only reciprocity and retaliation, it is clear that compliance can often be sustained by international law. But as discussed, neither of these strategies works especially well when the relevant issues involve public goods. Areas in which this is likely to be a problem include, for example, multilateral environmental agreements (where the harms from noncompliance are truly felt by many states), human rights agreements, and many multilateral arms control agreements. In each of these areas, threats of reciprocal withdrawal of compliance and threats of retaliation will normally lack credibility. Even for this limited subset

of agreements, however, there is no reason to conclude that international law is ineffective or cannot generate cooperation.

The discussion of reputation already presented, along with the detailed presentation in the next chapter, gives us a theory capable of explaining the multilateral cooperation that we observe, even in these public goods areas. Reputation can provide an incentive to comply with international obligations even when reciprocity and retaliation do not, because reputational sanctions require neither that states choose to impose costly sanctions in an effort to generate future compliance nor that reciprocal withdrawal of concessions is practical. Reputational sanctions, instead, reflect the updating of beliefs by self-interested states. There is no need for coordination, no need for formal adjudication of a dispute (though this can improve the effectiveness of reputational sanctions), and no need for costly actions by sanctioning states.