

# 3

## REPUTATION

The basic way retaliation and reciprocity work can be well understood with the brief discussion devoted to each in the prior chapter. Reputation, however, requires elaboration. This chapter provides a more complete picture of how reputation affects state behavior and how it interacts with other issues.

Just about every sort of cooperation that is interesting to us relies, in some way, on reputation. The primary exceptions are coordination games and other similarly straightforward games, discussed in chapter 2, in which compliance can be achieved without resort to reputation and, indeed, without resort to international law. The large number of interactions and games that remain all involve reputation in some way. It is true that cooperation can often be sustained through reciprocity or credible threats of retaliation, but as discussed in chapter 2, each of these relies on some form of reputation. Reciprocity can sometimes succeed because when a state has breached an obligation, its promises of future compliance lack credibility. The state suffers a reputational loss, at least with respect to the particular agreement at issue, and so the state's partner no longer has reason to believe promises of future cooperation. A threat to retaliate is costly to the sanctioning state, and so rational states will only take retaliatory action if it helps them to establish a reputation for punishing those that breach their commitments.

Reputation is also critical to our understanding of multilateral agreements. Where the benefits from such agreements are nonexcludable, reciprocity works poorly. It normally makes little sense for all states to cease their own compliance in reaction to a violation by only one state, and threats to do so lack credibility for that reason. Retaliation in multilateral agreements often suffers from significant free-rider problems, as

no state has optimal incentives to bear the costs of retaliation when all parties stand to enjoy the compliance benefits of those actions.

Reputation, on the other hand, can work effectively in the multilateral context. Indeed, it may work better in multilateral settings than in bilateral ones. Parties to an agreement may learn about violations more quickly and accurately than those outside of the agreement, and they may have a more finely tuned sense of relevant nonreputational payoffs. If so, the reputational consequences of a violation will be more severe in a multilateral context—because the reputational information spreads quickly to more countries.

This book is certainly not the first effort to use reputation to explain state behavior, and this chapter is not the first to address how reputation works (Abreu and Gul 2000; Alt, Calvert, and Humes 1988; Downs and Jones 2002; Keohane 1984; Maggi 1999; Mercer 1996; Sobel 1985). This chapter differs from prior writings in two important respects. First, it focuses on international law. Much of the existing writing on reputation addresses a state's reputation in the security arena.<sup>1</sup> This distinction is important because security-related interactions are only a small fraction of the interactions between states. Furthermore, security issues are aberrational because of the enormously high value states place on issues of national security and national survival. The importance of security issues has attracted scholars, but an excessive focus on high-value interactions fails to capture the logic of the more routine, and considerably more frequent, aspects of international legal relationships. Second, this chapter considers reputation in more detail than most previous studies and develops a more complete theory than has been done in previous work.<sup>2</sup>

The chapter first develops a simple model of reputational gain and loss and considers how state behavior is affected as a result. I then examine how relaxing certain informational assumptions affects the role of reputation. In particular, I point out that uncertainty about payoffs, legal rules, and actions taken all reduce the impact of reputation on behavior. I then discuss the extent to which states are likely to have a single reputation for compliance with international law as compared to multiple reputations. Finally, some of the limits of reputation's ability to affect behavior are discussed.

## How Reputation Is Gained and Lost

### *The Compliance Decision*

We begin by considering how reputation is acquired and lost. The theory of reputation is relatively new in the international law literature but is more established in economic and political science writings. Despite its use in several disciplines, however, we lack a comprehensive theory of reputation that can be applied in the international law context. The first task of this chapter, then, is to outline such a theory.

Chapter 2 defined reputation as judgments about an actor's past behavior used to predict future behavior. Consistent with that definition, we can define a state's reputation for compliance with international law as judgments about an actor's past response to international legal obligations used to predict future compliance with such obligations. This reputation is an estimate of the state's true willingness to comply even when nonreputational payoffs favor violation. This willingness to comply depends on the state's discount rate; the domestic politics in the state (e.g., the extent to which domestic political structures make violation of international law difficult or costly); the state's willingness to impose costs on others;<sup>3</sup> the value of future opportunities to cooperate (which a current violation may jeopardize); and so on.

Other states are assumed to be unable to observe this underlying willingness to comply, and so they must estimate it on the basis of the actions of the state. In principle, every observing state has its own perception of a particular state's reputation. Thus, the United States may have different reputations in Canada, Argentina, Russia, and Syria. For the moment, I abstract away from this issue and assume that every observer has the same view of the state's reputation. This assumption is relaxed later in the chapter.

In the prior chapter, I treated the acquisition and loss of reputation in an extremely simple way. As described there, states that honor their commitments acquire reputational capital, and states that violate their commitments lose it. A moment's thought, however, makes it clear that things must be more complicated than this. If it were simply a matter of counting the instances of compliant behavior, states could build their reputations by signing many treaties that impose trivial obligations. A sensible model of reputation building cannot, for example, lead to the conclusion that Bolivia, a landlocked country, can improve its reputation

by committing to keep its ports open. Similarly, it cannot be that the tiny island republic of Vanuatu, whose total gross domestic product is less than \$350 million, can improve its reputation by agreeing to refrain from placing weapons in space. The acquisition of reputation clearly must be more complex than simply complying with commitments.

The loss of reputation, similarly, must depend on more than simply how often a state violates its commitments. A minor, technical reporting violation that is quickly corrected, for example, must have less of an impact on a state's reputation than a large-scale and public reputation of a major treaty.

When entering into an agreement, states want their promises to be credible, and they must ultimately rely on reputation for that credibility. As the expected costs of performance increase, states require more credibility and, therefore, a stronger reputation for the associated promises to be believed.

An increase in reputation,  $Q$ , has value to the state in the sense that a higher reputation allows it to make more credible promises to other states and to extract more gains from its international engagements. To see how reputation can change, consider the following illustration.

Suppose a state has an existing reputation it has acquired as a result of its past conduct or, if it is a new state, it has some baseline reputation that represents the expectations of others regarding the state's willingness to comply with international legal obligations. Faced with a compliance decision, the state either complies with or violates its obligation, and this decision may impact the state's reputation.

It is assumed that other states are unable to observe the state's true willingness to comply with its obligation. Other states can, however, observe the acting state's nonreputational payoffs. They may know, for example, what it would cost the state to comply with an environmental obligation, but they do not know what the state hopes to achieve in the future through international environmental agreements or the state's discount rate. All states also know the legal rules in place and are able to accurately observe the behavior of the acting state. These issues (nonreputational payoffs, legal rules, and the ability to observe a state's actions) may bear on how state behavior affects reputation, and they are discussed later in this chapter, but the presentation is simpler if they are omitted for the moment.

The state must choose between compliance and violation. If the state chooses to comply, it receives a total payoff that consists of the nonrepu-

reputational payoff plus the reputational payoff. The reputational payoff measures the value of any increase in reputation generated by the decision to comply. If both parties to an agreement comply in every period, the payoffs corresponds to the comply-comply payoff in table 3 in chapter 2.

If, on the other hand, the state violates the legal rule, it receives the associated nonreputational and reputational payoffs. Now the reputational payoff reflects any loss of value to the state as a result of its violative conduct.

In deciding how to act, the state compares the total payoff in the event of a violation to what it would receive should it comply. Violation is only tempting to the state if the nonreputational payoff from violation is larger than the nonreputational payoff from compliance. The question, then, is whether the reputational payoffs can provide some offsetting incentive to comply.

In my later discussion of the way reputation changes in response to either compliance or violation, I show that a state's reputation will be changed only to the extent that the state's behavior differs from what observing states have expected. It is possible, therefore, for a compliant state to enjoy no reputational gain or for a violating state to suffer no reputational loss. But when choosing between violation and compliance, at least one of those actions will generate a change in reputation, because at least one of those actions will be different from what observing states expect. This means that the reputational payoff from compliance will be larger than the reputational payoff from violation. This difference in the reputational payoffs has the potential to offset the gains a state gets from violating an agreement.

Specifically, the state will comply if the reputational gain from compliance exceeds the increase in nonreputational payoffs available if it violates its commitment.

Returning to the question of how reputation is gained and lost, notice that the value of reputation will not be the same for every state or in every issue area. Some states (or states in some situations) are in a better position to extract value from a good reputation. These states might, for example, have many future opportunities for cooperation that require them to make credible promises, or they may have an ongoing relationship with a partner that makes a good reputation especially valuable. Reputation is more valuable for such states than it is for those with fewer

or less valuable potential international engagements. When reputation is more valuable, states are willing to bear more costs in order to enhance or protect it. They will, therefore, be more willing to comply with an international legal commitment.

To illustrate, suppose that the value of reputation is measured by a function,  $V$ . For simplicity, assume that there are two different types of states. One type can extract a great deal of value from any given level of reputation and faces a value function labeled  $V^I$ . The other type gets less value from a good reputation and faces a value function labeled  $V^L$ .

States facing the value function  $V^I$  have more to gain from a strong reputation than those facing the function  $V^L$ . Faced with a compliance decision, a state will take into account how changes to its reputation will affect its overall payoffs. As figure 3.1 indicates, a given change in reputation,  $Q$ , has a greater impact on the state that places a high value on reputation than on the state that places a low value on it. It follows that the latter will invest less in reputation (i.e., will violate obligations more readily) than will the former. In equilibrium, then, we expect to find some states (or some state-subject area combinations) with good reputations and others with weaker reputations.

As the value of reputation increases, a state is better able to resist the temptation to violate a legal obligation, meaning that the difference in nonreputational payoff between violation and compliance must be larger to provoke a breach by a state for which a good reputation is valuable. One implication is that we cannot hope to predict compliance behavior

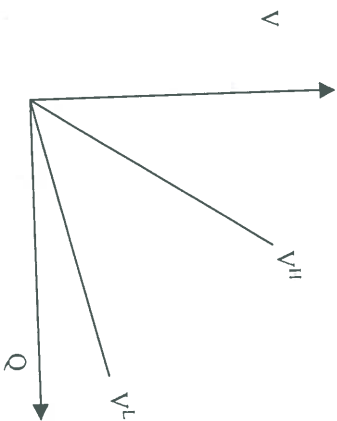


Figure 3.1 The Value of Reputation

without at least some understanding of both the reputational and non-reputational payoffs at issue.

This model of reputation has some similarities to the model developed in Mercer (1996), which focuses on a state's reputation for resolve, defined as its willingness to risk war to achieve its objectives. What he terms "situational" attributes correspond roughly with what I term "nonreputational" payoffs, and what he terms "dispositional" attributes correspond to my "reputational" payoffs. However, unlike Mercer, who takes an approach rooted in social psychology, I assume that states are rational actors. They evaluate the behavior of all other states, whether friend or foe, by using the available evidence and rationally attributing behavior to reputational or nonreputational payoffs on the basis of the information the observer possesses. States in Mercer's model also rely on past behavior, but do so in an asymmetric way. Specifically, he assumes that observers attribute undesirable behavior (from the observer's perspective) by their adversaries to reputation (to use my vocabulary), but attribute good behavior by the same states to nonreputational payoffs. His assumptions about allies are the opposite. With reference to a reputation for resolve, for example, the consequence of his assumptions is that "while adversaries can get reputations for having resolve, they rarely get reputations for lacking resolve; and while allies can get reputations for lacking resolve, they rarely get reputations for having resolve" (Mercer 1999, p. 10). Mercer also understates the relationship between reputational and nonreputational issues. He argues that observers attribute state actions to either reputational issues or nonreputational issues and asserts that a "situational (i.e., nonreputational) attribution cannot sensibly be used to predict behavior in a different situation" (p. 15). When it is recognized that reputational and nonreputational payoffs both influence compliance decisions, however, it is clear that a compliance decision taken in one situation yields information about how a state will behave in other situations. This information, in turn, can be used to predict future behavior.

The next step in developing a theory of reputation is to consider in more detail the circumstances in which reputation is likely to change. There are three key factors likely to determine whether a particular action affects the associated reputation.<sup>4</sup> These are (1) the nonreputational payoffs a state is facing; (2) the state's existing reputation at the time of the action; and (3) the importance of the obligations to other states. These factors help to explain why a decision to violate or comply

will have different reputational consequences in different contexts. They also address one of the key arguments advanced by international law skeptics attempting to dismiss reputation.<sup>5</sup> Because reputational acts at the margin and nonreputational factors also influence state decisions, it should come as no surprise that different treaties (or customary international law) generate different levels of compliance at different times. Reputation has an impact on the compliance decision by making compliance somewhat more likely, not by ensuring the same rate of compliance in all circumstances.

#### *Compliance Decisions and Their Effect on Reputation*

Because gains and losses of reputation must depend on more than simply whether agreements are complied with or violated an agreement. In fact, the about why a state has complied with or violated an agreement. In fact, the reasons for a violation may be as important as the actions itself. In our model, a state acts in an effort to maximize its payoffs, and those payoffs are divided into two categories: reputational payoffs and nonreputational payoffs. Consider first the nonreputational payoffs.

A state that has a powerful national security reason to violate an agreement, for example, faces nonreputational payoffs that provide a strong incentive to ignore the commitment. In contrast, a state whose behavior would be consistent with the agreement even in the absence of a commitment faces nonreputational payoffs that give it an incentive to comply. The reputational consequences of an action (i.e., the reputational payoffs) are closely related to these nonreputational payoffs.

When entering into an agreement, states hope that both sides will comply, but they also recognize that a violation may occur. In fact, they recognize that under certain circumstances, violation is to be expected. Countries that have signed the Kyoto Protocol, for example, have agreed to limit their "carbon dioxide equivalent emissions" of certain greenhouse gases. A failure to do so would represent a breach of the commitment made in that protocol and would normally lead to some level of reputational sanction, but the magnitude of that sanction depends on the context. A violation of the protocol that is plausibly attributable to the fact that the country is at war and devotes its efforts to the pursuit of the war effort rather than compliance with emissions rules will generate a smaller reputational sanction than a violation that cannot be justified in some similar way. Though the protocol does not provide an

exception for national emergencies, countries recognize when they sign any agreement that there are circumstances in which compliance will not be forthcoming. Thus, article 18 of the Kyoto Protocol provides that when addressing cases of noncompliance, the governing conference shall develop "an indicative list of consequences, taking into account the cause, type, degree, and frequency of noncompliance." As this section contemplates, there are situations in which compliance is not to be expected. It follows that reputational sanctions will be quite modest in those circumstances. To see why this is so, imagine the position of a state negotiating an environmental agreement. Compliance with the agreement will impose a cost on the signatories, but all parties prefer mutual compliance to mutual noncompliance. The signatories expect compliance in many states of the world, but not in every such state. For instance, assume that every signatory recognizes that a country will abandon its obligation if it goes to war because the environmental obligations are simply too costly to accept during wartime. If military conflict or severe domestic unrest explains why a potential signatory breached a similar obligation in the past, will this fact hamper its ability to participate today? As long as all parties expect breach in the event of a war, there is no reason that past conduct consistent with this expectation would affect the negotiation. In effect, the agreement has an implicit exception in the event of war. If war was the only concern, the parties could, of course, include it explicitly. But if the set of circumstances in which states recognize that compliance is unlikely includes a wide array of reasons for noncompliance, it becomes impractical (or impossible) to specify those reasons in the agreement.

To illustrate more clearly how nonreputational payoffs might affect reputation, consider the NPT.<sup>6</sup> The main provisions of the NPT require that state parties designated as non-nuclear weapons states (NNWS), a category that includes all but five countries (the United States, Great Britain, France, China, and Russia as the successor to the Soviet Union), refrain from seeking nuclear weapons, and that all states refrain from aiding NNWS in acquiring nuclear weapons. Currently, the treaty has 187 member states. For many, if not most, of these states, compliance with the treaty comes as no surprise. The technological expertise and the industrial infrastructure necessary to support a nuclear weapons program would require such a major investment over so many years that a simple cost-benefit analysis dictates compliance for most countries. It is difficult to imagine Trinidad and Tobago starting a nuclear program,

even if it were not a treaty member. For other states, domestic politics make a nuclear program quite unlikely. In Germany, for example, there is strong public opposition to any radical expansion of military forces, and there would certainly be strong objection to a proposed nuclear weapons program.

The designated nuclear weapons states (NWS), on the other hand, are not constrained by the treaty's first provision (the prohibition on acquisition of nuclear weapons) and so cannot violate it. For at least some of the NWS, including the United States, violation of the second provision (prohibiting states from aiding NNWS efforts to acquire nuclear weapons) is unlikely, both because the proliferation of nuclear weapons technology would destabilize the community of nations and because it would undermine the NWS' existing nuclear and conventional military superiority over NNWS.

So for many NNWS and NWS, one would expect compliance even in the absence of a treaty. Because the nonreputational payoffs suggest that these states would behave consistently with the treaty obligations even if the treaty were not in existence, observer states have no reason to interpret compliance as any sort of positive signal about reputation. For these states, compliance provides no reputational gain relative to a world in which the treaty did not exist.

Some of the NNWS, however, have an incentive to pursue nuclear weapons. South Africa, for example, had a nuclear weapons program beginning in the mid- to late 1960s.<sup>7</sup> In fact, the program was pursued to completion and produced a weapon in 1982. Within a decade, however, South Africa had dismantled its program and publicly renounced its former nuclear status. Though it is true that South Africa's nuclear deterrent had lost much of its value with the end of the Cold War and the cease-fire leading to the end of the Soviet-backed war in Angola, possession of nuclear weapons would have continued to be of value. Nevertheless, because South Africa in the early 1990s was attempting to reform its image as a global pariah, reputational considerations (along with other factors) dictated that it bring itself into compliance with international norms. It had to choose between its nuclear weapons and participation in the NPT, because the NPT does not provide for an increase in the number of nuclear weapons states beyond the five that qualified by virtue of having tested a device prior to January 1, 1968. In

1991, one year after it began dismantling its nuclear weapons program, South Africa acceded to the treaty. Why would South African want to join the treaty? The obvious answer is that doing so allowed it to more fully rejoin the community of nations and reap the benefits of cooperation with other states. It sought to improve its reputation.

Needless to say, these examples might be contested. These events, like those in any of the examples used in the book, were affected by other forces, and it is possible that these other forces offer a better explanation of events. As I discussed in the introduction, examples such as these are only intended to illustrate the points being made. In this case, the point is that some states that appear to have an incentive to pursue nuclear weapons have chosen to refrain from doing so, and the presence of the NPT may well have encouraged that decision.

Returning to the NPT example, there are states that not only have strong incentives to develop nuclear and other nonconventional technologies (or to traffic in them) but have gone ahead and done so. For some states, this behavior is expected because they have an existing reputation that is sufficiently low that they are not actually expected to comply, and they may have nonreputational payoffs that provide a strong incentive to violate the agreement. This is plausibly the case for Iran and North Korea (especially the latter), both of which were parties to the treaty prior to North Korea's withdrawal in 2003. Already pariah states, these countries have little reputational capital, so other states likely did not expect membership in the treaty to be a meaningful deterrent. Though they could have improved their poor reputations through compliance (or by dismantling existing programs), such a strategy promised no more than modest gains. Rebuilding a reputation can be a slow and costly process that may not suit the needs of these states. Furthermore, it is unlikely that these states would have developed extensive and important cooperative arrangements to which a better reputation would have added value. Whatever reputational gains they might expect are too small to offset the incentive they have to develop nuclear weapons.

Another possibility, illustrated more clearly by the Russian experience, is that the payoffs from noncompliance are too great, even for a state that has a lot to gain from an improved reputation. Russia, a party to the NPT, agreed to complete the construction of first one, and then possibly a second, nuclear reactor for Iran at Bushehr. The deal was financed in large part by loans from Russia to Iran, so failing to complete

the project would have had, among other things, an opportunity cost in the form of a loss of exports for the cash-strapped Russian state.

As Russia emerged from communism, however, it continued to have an incentive to engage the rest of the world and to enter into a broad variety of international agreements. With concern about Iran's nuclear ambitions mounting in the West, the United States tried to persuade Russia to cease cooperating with Iran, linking a cessation of cooperation with Iran to future cooperation with the United States. This U.S. strategy brought reputation into play explicitly, as well as introducing a threat of retaliation. Despite this and other potential costs, however, Russia continued to cooperate with Iran for several years. The potential payoff from dealing with Iran was simply too large to be offset by either the harm to Russia's reputation or the threat of sanctions by the United States. Eventually, Russia began to express misgivings about Iran's intentions and capabilities at Bushehr. While the United States and Europe have been the most critical of the Iranian nuclear program, Russia has steered a middle course, offering to enrich uranium for use in the Iranian reactors, thus denying Iran the enrichment technology essential for making a uranium bomb. Russia also stalled before eventually voting on the International Atomic Energy Agency's governing board to report Iran to the UN Security Council for possible penalties stemming from its nuclear program. These compromise actions likely reflect Russia's desire to balance its reputational risk against its nonreputational incentives to cooperate with Iran on nuclear technology. In other words, Russia does not wish to forgo the benefits, economic and otherwise, of cooperating with Iran, but Russian leaders are aware that if other states perceive Russia as aiding Iranian efforts to obtain a nuclear bomb, Russia will suffer greater reputational harm. Russia's hesitation as to whether to continue to assist Iran indicates that the issue is a close call, and thus one in which international law can influence state decisions.

The NPT example demonstrates how a single international legal obligation interacts with the payoffs of states. For some states the obligation will have no effect, because they would have behaved consistently in any case. For other states behavior is not changed, because the treaty is insufficient to prevent a violation. For a third group, however, the obligation can change behavior, because the reputational consequences of a violation can be enough to tip the scales in favor of compliance.

### *The Role of Nonreputation Payoff and Existing Reputations*

Sticking with the NPT example, I now turn to a more thorough examination of how reputation and payoffs interact. For concreteness, I focus on the Russian experience with the NPT. Other nations' expectations of Russian behavior were not the same as their expectations of, for example, Great Britain or the United States. First, Russia's economic circumstances are considerably different from those of the Western nuclear powers. The incentive to export nuclear technology for economic profit would be considerably higher for Russia than for a more financially secure Western nation. These incentives, and the fact that they are known to other states, leads to a reduced expectation of strict compliance with the NPT.

Second, Russia's preexisting reputation for adherence to the NPT was likely not glittering. In other words, other states would think it more likely that Russia, compared, again, to a Western nuclear power such as Great Britain, would undertake an unauthorized transfer of nuclear technology. This preexisting reputation has the same effect as public knowledge of incentives to not comply; it means that if Russia were to transfer technology to a nation with nuclear ambitions, Russia's reputation will not suffer to the same degree that Great Britain's would if it did the same thing. Thus, perhaps counterintuitively, the fact that everyone believes Russia is more likely to violate the letter or spirit of the NPT actually reduces the reputational sanction when such a violation occurs.

Assuming that Russia behaved consistently with the expectations of other states, no change in the beliefs of those states was warranted. This could be so, for example, if Russia faced payoffs that so forcefully called on it to violate the NPT that existing beliefs about Russia and its reputation led observing states to expect violation.

More generally, in order to determine if a state's actions will affect its reputation, it is necessary to know something about both reputational and nonreputational payoffs. When the nonreputational payoffs generate a sufficiently strong incentive to violate an agreement, a decision to breach may not lead to any change in reputation. Similarly, if a state's nonreputational payoffs provide an incentive to comply with an agreement, the decision to comply will not lead to a reputational gain.

In addition to the nonreputational payoffs, the impact of an action on reputation depends on that state's existing reputation. Consider, for example, a state that complies with a treaty, despite nonreputational

payoffs that counsel violation. If that state has a very strong reputation prior to its decision, the decision to comply will reaffirm that good reputation, but may not increase it. The decision to comply may still offer a benefit, however, by solidifying the state's reputation and making the observing states' beliefs more stable going forward.

On the other hand, if the state started with a poor reputation, the decision to comply would cause observing states to adjust upward their estimate of the state's willingness to comply. The analysis is the same for states that violate the treaty, but the impact is reversed. Suppose a state's nonreputational payoffs suggest violation, but its reputation is strong enough that observing states anticipate compliance. If the state violates the agreement, other states will revise downward their estimate of the violating state's willingness to comply. In contrast, a violating state with little reputation to start with may not face any change in its reputation because its actions simply confirm existing perceptions.

So we can predict that a state's decision to comply with a legal rule will enhance its reputation when the nonreputational payoffs counsel violation and the state's existing reputation (as other states perceived it) is insufficient to cause observing states to expect compliance. A violation will harm a state's reputation when the nonreputational payoffs, combined with the state's existing reputation, predict compliance.

The foregoing is really just a claim that states engage in Bayesian updating of their estimates of other states' willingness to comply with international law.<sup>8</sup> The reputation of a given state is determined by that estimate, which can change over time. Notice that because states update their beliefs in Bayesian fashion, the strength of their *ex ante* beliefs matters. For a given decision to violate or comply, a state with a long established and consistent reputation will suffer a smaller reputational sanction than one with a less established reputation. Other states will have stronger priors about the former state than about the latter, and so the former's reputation will be less affected by an individual action.

### *The Importance of the Obligation*

One would like, of course, to know not only whether a state's actions have reputational consequences but also the magnitude of those consequences. Because this book takes a theoretical approach, however, there is no way

to provide an estimate of the magnitude of reputational sanctions. The theory cannot specify (without additional assumptions) the importance of reputation to states either in absolute terms or relative to nonreputational payoffs. It is possible, however, to consider some of the factors likely to affect the size of the reputational sanctions a state will face if it violates an obligation (or, conversely, the magnitude of the reputational benefit if it complies). I will discuss more of these factors later in this chapter, here I want to point out the role played by the importance of the obligation.

The point here is fairly obvious, but nevertheless bears mentioning. The relative importance of an international legal obligation affects the reputational consequences of violating it. Thus, a refusal to allow inspection of nuclear reactors under the NPT represents a serious breach of a state's commitments and will generate a strong reputational reaction. In contrast, minor violations, such as missing a reporting deadline, are unlikely to be viewed as a serious breach and, therefore, are unlikely to lead to significant reputational sanctions. One can think of this measure of relative importance as a measure of the extent to which the interests of other states have been affected. Refusing to allow inspectors onto the site of a nuclear reactor undermines the very purpose of the NPT and compromises the interests of other parties to the agreement. It is true that these other states have not been directly harmed, but the inability to monitor compliance with a treaty of such importance, combined with the negative signal that a refusal to allow inspectors entry sends, raises significant security concerns in other states.

If violations of more important obligations lead to larger reputational sanctions, one might conclude that such agreements are more likely to be honored. This, however, does not follow, and in fact, the opposite is likely to be true. The most important obligations are those where cooperation yields large gains. They are also the ones where the incentive to defect is largest. To say that an agreement yields large gains from cooperation is equivalent to saying that cooperation would be difficult without the agreement—meaning that one or more parties has strong incentives to defect. In such "high stakes" commitments, we should expect both nonreputational and reputational payoffs to be large.

Though both these payoffs grow with the stakes, it seems likely that at least with respect to agreements that implicate the highest stakes (e.g., security), reputational sanctions will normally play a (perhaps vanishingly) small role. Reputational sanctions, after all, can only do so much.

Even a complete loss of reputation is something a state can recover from, and so there is some upper bound on the magnitude of reputational sanctions. As the nonreputational payoffs of a decision increase, the likelihood that they will swamp the reputational aspect of the decision increases. Thus, for example, one should not expect legal commitments regarding the conduct of war to be especially effective. To the extent that we see compliance in such cases, it is likely to be explained by nonreputational rather than reputational payoffs. Furthermore, as Morrow (2000) notes, major future challenges to a state's security or even its existence may be difficult to foresee at the time the decision to comply with an alliance agreement is made. Because such situations arise infrequently or unpredictably, the reputational costs to violating alliance agreements may be negligible (because a reputation for complying in situations that do not frequently occur is worth less than a reputation for complying in more common situations) when compared with the substantial costs of fighting a war.

#### Managing Reputation over Time

The reputation of a state is contingent on its past behavior, but it is the knowledge that today's conduct will affect tomorrow's reputation that gives reputational sanctions their force. When making any single decision about compliance, the impact of that decision on a state's reputation and future payoffs is taken into account. Because states recognize the incremental effect of compliance decisions on reputation, they are able to make strategic decisions about how to manage their reputation over time. For example, when recent conduct has damaged a state's reputation, the state may set out to rebuild it by complying with international legal rules it otherwise would have been tempted to ignore. More generally, the process by which states accumulate and erode their reputations should itself be viewed as the result of a cost-benefit analysis. Because different states (or individual states at different times) will face different costs and benefits, they will differ in their willingness to build and maintain strong reputations. For example, a state whose reputation is sufficiently tarnished may find it so costly to repair that it is better off extracting everything it can in the short term and ignoring much of international law. North Korea might fit this category. For such

countries, reputational constraints have little or no force, and international law can only play a role to the extent it generates direct sanctions, is made effective through reciprocity, or assists in the management of simple problems of cooperation such as coordination games.

By way of example, consider Libya's role in the bombing of Pan Am Flight 103 over Lockerbie, Scotland, in 1988. This bombing was carried out by Libyan agents, and as such constitutes an "arbitrary deprivation of life," in violation of the UN Covenant on Civil and Political Rights, to which Libya is a party. Libya, at the time, was arguably a pariah state, at least in the eyes of Western states, and so had little reason to be concerned about its international reputation. (We are talking here about a reputation for compliance with international law. Libya's actions implicated other forms of reputation, of course. Later in this chapter I discuss how these different types of reputations interact.) A decision to violate this treaty was unlikely, for example, to significantly hamper Libya's future ability to benefit from international cooperation, because such cooperation was unlikely to come about in any event. So Libya had little reason to resist whatever short-term benefits it gained from its illegal actions. It was only after years of direct sanctions by the United States and the United Nations that Libya agreed, in 1999, to hand over for trial the agents responsible for the bombing. By delivering the two agents accused of the bombing (one of whom was convicted) to Scottish authorities, Libya signaled that it was prepared to begin complying with international law. This action, along with an agreement to pay damages to the families of the victims, played a role in Libya's ongoing reintegration into the international system.

In other contexts, a state may choose to violate international legal norms in one sphere while attempting to bolster its reputation in another. This is arguably what the South African government did during certain periods of the apartheid regime. While that government was oppressing its nonwhite populations in contravention of the international law of the time, it was otherwise "scrupulous in its observance of international law" (D'Amato 1971, p. 30).

These examples demonstrate that states will not always seek to preserve or build their reputations. Reputation, like any other asset, can not only be acquired but also used or "spent." States will sometimes choose to extract whatever short-term gains they can, despite the fact that doing so will erode their reputation. This incentive to defect from

international rules will come about when the gains from violation are especially high or the loss to reputation from breach is especially low. Because a state's reputation may be subject-matter specific (i.e., a state may have different reputations over different issues) a decision to be- have opportunistically and extract short-term gains may be profitable to states when the long-term value of cooperation in an area is modest.

Consider an example in the foreign investment area. In the 1960s and 1970s, many newly independent states chose to expropriate the as- sets of foreign investors. These actions were arguably contrary to cus- tomary international law, at least early on.<sup>9</sup> The decision to expropriate can be seen as a decision to extract current value at the expense of future reputation. The states involved had the opportunity to capture rents from existing stocks of foreign investment, especially because much of the investment consisted of fixed capital that could not be removed. These states were also attempting to establish their economic indepen- dence and often adopting policies that were hostile to new flows of for- eign investment. The knowledge that their policies already made large future inflows of foreign investment unlikely made the decision to ex- propriate much less costly. Put another way, compliance with what many Western states believed was a rule of customary international law (a prohibition on expropriation without prompt, adequate, and effec- tive compensation) was of modest value, because a reputation for com- pliance in the investment area was unlikely to generate a large increase in investment at the margin.

A great deal of foreign investment into developing countries is now governed by bilateral investment treaties. These agreements explicitly for- bid expropriation except for a public purpose and (typically) with prompt, adequate, and effective compensation. They also provide for mandatory arbitration in the event of a dispute between a foreign investor and a host state. By entering into one of these treaties, a host state makes a reputa- tional commitment and, as a result, has an enhanced incentive to refrain from expropriating or otherwise treating investors in a manner that is contrary to the relevant treaty. Notice that the incentive to comply comes both from the fact that the relevant rules now take the form of formal treaties and from the fact that the value of a reputation in this area is higher than it was 30 or 40 years ago. Developing countries today are more likely to consider foreign investment to be part of their development strategy for the future. A loss of reputation in this area would reduce the future flow of foreign investment, which these states perceive as costly.

Consider one more example. Following the Russian Revolution of 1918, the new Bolshevik government repudiated the debts accumulated by the tsarist government, claiming, among other things, that they were the personal debts of the tsar. This action was a clear violation of Russia's international legal obligations.<sup>10</sup> The Bolsheviks, however, had no reason to invest in their reputation for compliance with international legal rules. In the early days of the regime, they expected a communist revolution to overwhelm Europe in the wake of World War I. Such a revolution would make past reputations irrelevant, as a new era of in- ternational relations would be born. This view was so ingrained in the Soviet leadership that Leon Trotsky, the first Soviet foreign minister, defined his role as being merely to "issue a few revolutionary procla- mations to the peoples of the world and then shut up shop" (Carr 1985, p. 16). The potential gains from cooperation, then, were perceived to be modest, making repudiation of the debt more attractive. Even if Trotsky's words were mere rhetoric and the Soviet leadership in fact anticipated ongoing relations with the West, the repudiation of the debt is easy to understand. The new Bolshevik government was sufficiently hostile to the West that it had no reason to expect a continuing flow of loans and so no reason to be concerned about whether promises to repay loans would be credible. Whatever reputational harm it incurred by repudiating the debt was of little consequence because, at least with respect to financial matters, the Soviet Union had nothing to lose.

In the late 1980s and early 1990s, the Soviet Union and (after its collapse) Russia were attempting to do just the opposite of what the Bolsheviks had done; they were going to considerable lengths to rebuild a reputation for compliance with international legal obligations. Fol- lowing the collapse of the Soviet Union in 1991, the new Russian gov- ernment announced that it would honor existing Soviet debt, despite the change in regime. Had it chosen to repudiate the debt, the new Russia would have compromised its reputation and undermined its efforts to develop financial ties with the West. Abiding by international norms helped Russia to build a new reputation that, in turn, increased its ability to attract financial assistance from the West.

This example illustrates another feature of state reputations. Older, more stable states will have more established reputations. Each instance of compliance or violation will have less impact on the reputation of such states because other states' prior beliefs will be stronger. Once a state is established, improving its reputation, as Mikhail Gorbachev, the

Soviet President, sought to do in the final days of the USSR, can be a difficult and expensive proposition.

Political scientists have used signaling theory to model the cost to a state of demonstrating that it is trustworthy. For example, Kydd (2000a) notes that if state A wishes to engage state B in a cooperative venture, but state B distrusts state A, costly signals will be necessary to reassure state B that state A is a reliable partner. In the case of Gorbachev, Kydd analyzes the Soviet capitulation over the Intermediate Range Nuclear Forces (INF) Treaty in 1987 as just such a costly signal. That treaty called for the United States and Soviet Union to eliminate all ballistic and cruise missiles with ranges of 500–5,500 kilometers and provided for verification procedures, including on-site inspections. This treaty represented a significant concession by the Soviets for a variety of reasons, including the fact that at the time the verification procedures in the treaty were the most stringent in history, including intrusive on-site inspections previously opposed by the Soviets; the category of missiles in question gave the Soviet Union many more targets (e.g., American allies in western Europe) than it gave the United States; and the treaty did not govern the disposition of the nuclear weapons of other Western nations, meaning that American allies remained free to deploy intermediate-range missiles in Europe if they so chose.

For a signal to work, it must be cheap enough for the sender to decide to proceed but expensive enough that the target, in this case the United States, updates its prior beliefs about the reliability of its prospective partner. In reference to reputation more generally, we should expect the cost of the signal necessary to cause such an updating of priors to be higher when the sender is an old state with a well-established reputation, or when states A and B have a long history of interaction with a large number of data points supporting their prior beliefs. Because well-established reputations are less responsive to individual interactions, the magnitude of the interaction, in other words the cost of the signal, must be greater for it to have a significant reputational effect.

New states, or states with new regimes, on the other hand, have reputations that are less well established, so more is at stake when they make compliance decisions. Because observing states have only weak priors about the new state's willingness to comply with international legal obligations, each individual compliance decision has a larger impact on the new state's reputation. It follows that, all else equal, the incentive to

comply is increased. This may lead new states to be especially diligent about complying with international legal obligations (Shihata 1965), as is illustrated by the efforts of the Russian Federation (a new regime) to comply with its obligations in the early 1990s. The theory predicts, then, that states with more fragile reputations will make greater efforts to comply with international law than would a similarly situated state with a stronger reputation.<sup>11</sup>

Following the disintegration of the Soviet Union, several of the 15 new republics were "born nuclear," in the sense that they inherited an extensive nuclear arsenal from the Soviet Union. Global leaders hoped that Russia would succeed to the Soviet Union's status under the NPT as a nuclear nation and that the remaining states, Ukraine, Belarus, and Kazakhstan, would return their nuclear arsenals to Russia and accede to the NPT as NNWS. Considering Russia's history of imperialism with respect to its neighbors, it would have come as no surprise if those states had elected to retain their nuclear deterrent. Instead, all three complied, returning their weapons to Russia and acceding to the NPT as NNWS. A study of Ukraine's motivation for this action, taken in the face of realist incentives to the contrary, suggests that Ukraine was motivated by the desire to build its international reputation (Sagan 1996).<sup>12</sup>

Notice that this might be perceived to be inconsistent with the earlier discussion of newly independent states and their decision to expropriate investment. Closer consideration, however, reveals that there is no conflict here. As mentioned, newly independent states in the investment context had reason to think that future investment flows would be modest, so the expropriation of existing investment, even if it had a profound impact on the state's reputation for compliance with customary international law rules governing expropriation, generated modest costs.<sup>13</sup> The former Soviet Republics, on the other hand, generally sought relations with the West and could hope to extract value from a strong reputation.

### The Role of Information

UP to this point, it has been assumed that states have complete information about existing nonreputational payoffs, legal obligations, and one another's actions. We now relax those informational assumptions.

### *Uncertainty about Payoffs*

When states interact, they have some information about the goals and objectives of their counterparts but normally do not have complete knowledge of what they seek to achieve. There may, for instance, be domestic pressures that are not evident to outsiders. To model this fact, international objectives that are not known to outsiders. To model this fact, assume that states have imperfect information about one another's reputational and nonreputational payoffs. Rather than knowing these payoffs with precision, they have only unbiased estimates of them. Formally, one can think of the nonreputational payoffs in a particular instance as being drawn from a probability distribution where the observing states know the relevant distribution but do not observe the individual draw.

This imperfect information with respect to payoffs makes it more difficult to draw inferences from observations of compliance or violation. Suppose, for example, a state is observed violating a rule of international law despite the fact that other states expected it to comply on the basis of their estimates of its reputational and nonreputational payoffs.

An observing state might conclude that the nonreputational payoffs are just as it estimated, in which case its estimate of the other state's willingness to comply (i.e., its reputation) needs to be adjusted downward. Alternatively, the observing state may conclude that its estimate of the nonreputation payoff was incorrect. If the short-term payoff from the violation was larger, relative to the payoff for compliance, than the observing state expected, the violation can be explained. If so, the acting state may have behaved precisely as its existing reputation predicted, and so there is no reason to adjust beliefs about its willingness to comply.

In one circumstance, an observed violation is consistent with a state's existing reputation, whereas in the other, the violation signals that the state's reputation should be adjusted downward. Because both the willingness to comply with international law and the payoffs are uncertain, however, observing states have no way of knowing which of these two interpretations is correct (or if it is a combination of both). Lacking information, observing states must estimate the probability that the violation is due to unexpectedly high costs of compliance, and the probability that the state's reputation is weaker than was thought.

Uncertainty about the payoffs facing a counterparty, then, makes the process of estimating its reputation less accurate. Actions that should be attributed to nonreputational payoffs will be partially attributed to reputation, and actions that should be attributed to reputation will be partially attributed to other payoffs.

The lesson here is that uncertainty about nonreputational payoffs reduces the reputational consequences of a violation. These consequences are still present, but they are smaller. The same, of course, will be true of decisions to comply. Though they may increase a state's reputation, the magnitude of that increase will be reduced to the extent that observing states are uncertain about the associated nonreputational payoffs.<sup>14</sup>

### *Uncertainty about Legal Rules*

A similar problem arises if there is uncertainty about the legal rules in effect. Uncertainty of this sort may exist for any number of reasons. The rule at issue may be a rule of customary international law, in which case there is a good chance that its very existence is contested. Even when dealing with a treaty, there may be uncertainty because the treaty is ambiguous or incomplete with respect to the action in question. To illustrate, the U.S. government initially interpreted treaties pertaining to the treatment of captured al-Qaeda operatives and Taliban soldiers in ways that greatly enhance the freedom of American military and intelligence agencies to determine the manner in which they hold and interrogate prisoners. The United States contended, for example, that Common Article 3 of the Geneva Conventions, which provides certain rules governing conflicts "not of an international character," did not apply to the conflict with al-Qaeda and the Taliban. To the extent that these claims and the related protests from other governments and human rights organizations evidence some international uncertainty regarding the content of the relevant obligations, they reduce the reputational consequences to the United States for its alleged violation. In June 2006, the United States Supreme Court ruled that Common Article 3 does apply to individuals captured in Afghanistan, eliminating doubt about the meaning of Common Article 3, at least under U.S. law.

Faced with uncertainty as to the law, a state may have difficulty determining if it will face a reputational sanction for its conduct. Because the reputational sanction is the result of an updating of beliefs by

observers, what ultimately matters is the attitude of observing states. It is often clear to all that the law is ambiguous in some way, but different states may have different beliefs about what the law means or what it should mean. If the behavior of the acting state prompts observing states to adjust their estimates of the acting state's reputation downward, this represents a cost for the acting state.

This presents a dilemma for a state that is considering actions of uncertain legality. Even if the state determines, in good faith, that a particular action would be legal, it cannot be sure that other states would reach the same conclusion. To the extent that legal rules are uncertain, then, the acting state must tread cautiously.

An observing state has a related problem. If it observes conduct it considers illegal but the acting state argues is permitted, what conclusion should it draw about future compliance by the acting state? On the one hand, the acting state may consider itself to be in violation of the law and be attempting to muddy the waters by claiming that its actions are permitted. On the other hand, the state may have attempted in good faith to comply with the relevant rule and may simply have reached a different conclusion about its content. In either case, the reputation of the acting state may be affected, but the former would surely provoke a more significant reduction in reputation than would the latter. A good faith effort to comply indicates that the acting state sought to comply and its violation indicates a difference of interpretation. Had the contours of the law been clear, it may be that the state would have complied, in which case no reputational adjustment is called for. A state making a bad faith claim of compliance, on the other hand, knowingly violated the law and, in so doing, delivered a stronger signal about its willingness to do so.

The only sensible strategy for an observing state is to adjust its perception of the acting state's reputation on the basis of its estimate of that state's actions. That it violated a legal rule (as interpreted by the observing state) suggests that there will be some reputational loss (putting aside the possibility that the violation was expected on the basis of the nonreputational payoffs), and the magnitude of that loss will reflect some estimate of the likelihood that the acting state acted in good faith.

A closely related question is how states come to believe that a particular interpretation of a text is correct. That is, when there is ambiguity in an agreement, how do the states come to a view as to its meaning and determine what constitutes a violation? It seems clear that norms arise

around specific terms and that these norms influence interpretation. Less clear is where these norms come from. Some arise during the course of the negotiations, some from the general context (e.g., if a term is used in many existing agreements, it may be understood to mean the same thing in a new agreement), and some from interactions that take place after the treaty enters into force. Strictly speaking, of course, it is usually only the text itself that has legal force, but in practice, norms of interpretation are relevant.

But if norms serve to clarify the meaning of text, one must also acknowledge that norms can alter that meaning. In other words, norms may generate a *de facto* legal rule that the text of an agreement does not provide. Consider, for example, Trinidad and Tobago's withdrawal from the Inter-American Convention on Human Rights and the Optional Protocol to the ICCPR in 1999. Both of these withdrawals were facially legal, in that they were made pursuant to withdrawal provisions in both agreements.<sup>15</sup> Nevertheless, Trinidad and Tobago received an enormous amount of criticism for its withdrawal, which had the effect of denying supranational review of alleged domestic human rights abuses.<sup>19</sup> Denying this review was, in the critics' view, tantamount to denying the underlying substantive rights (Trinidad and Tobago did remain a party to the ICCPR itself). The overall effect was to undermine the perception of Trinidad and Tobago's commitment to its human rights obligations. One might interpret this as a signal that Trinidad and Tobago is unwilling to comply with customary international law rules embodied in the relevant instruments, or one might interpret it as a norm-driven *de facto* ban on withdrawal from the treaties themselves. In either case, despite the textual legality of its actions, Trinidad and Tobago's reputation for compliance with its human rights obligations suffered as a result of these withdrawals.

This same reasoning helps to explain how an agreement can come to be either partially or entirely obsolete. Even without formal termination, the relevant parties may all come to share the view that the treaty is inappropriate and no longer in effect. If that happens, a violation would have no reputational effect. Consider, for example, that the U.S. State Department lists as treaties in force a series of agreements the United States signed with Iran before the Iranian revolution of 1979, including a Mutual Defense Assistance Agreement,<sup>17</sup> economic cooperation agreements,<sup>18</sup> and educational assistance agreements.<sup>19</sup> It is clear that the United States' failure to come to the aid of Iran during the

Iraq war, and indeed its decision to assist Iraq, did not harm the United States' reputation for compliance with international law. Events had made the treaty obsolete, and all observers understood as much.

This discussion of information sheds light on one of the ways international agreements can affect behavior. By providing greater transparency with respect to the relevant rules, international agreements make it easier to classify behavior as compliant or violative. This, by itself, improves the probability of compliance because it distinguishes violators from nonviolators. Clarity in the rules makes it less likely that a violation will be perceived as compliant or that compliant behavior will be perceived as a violation.

This discussion also offers an explanation for why states go to such lengths to claim that their behavior is in compliance with international law. To the extent that a state is able to persuade others that its actions were actually in compliance with a reasonable interpretation of the law, the reputational sanction will be reduced.

#### *Uncertainty about Actions*

An analogous discussion applies to uncertainty about a state's actions. For example, it is not entirely clear how the United States is treating detainees in Guantanamo. There are credible allegations of human rights abuses, but uncertainty remains regarding the scale and scope of behavior that violates international law.

Needless to say, a violation of international law generates a reputational sanction only if some other country knows about the violation. It follows that a violation will lead to a smaller reputational loss if fewer countries know about it. By reducing the visibility of their violations, then, states reduce the reputational consequences.

Things are more complicated if a state's actions can only be observed imperfectly. This is true in the Guantanamo example, as well as a host of other contexts. For example, a state's environmental laws can be examined, but it may be difficult to evaluate the application of those laws or the degree of corruption associated with them. Other examples are easy to find—compliance with arms control agreements is often difficult to verify (e.g., Iran and its nuclear ambitions), the extent to which a state subsidizes an industry may not be clear (making it difficult to tell if it is violating its trading obligations), and so on.

In other instances, some states will observe a particular action but others will not. Those that observe the illegal action obviously do not face uncertainty with respect to the activity. If these observing states publicly announce that there has been a violation, however, those that did not observe the conduct may be uncertain about whether to believe the allegations or the denials of the accused state (assuming that the accused state makes such denials). These situations, and others like them, cause uncertainty about the actions taken by a state.

When there is uncertainty about whether a state has acted illegally, observing states must draw inferences from the information they have. They will estimate the probability that a violation has taken place and update their beliefs accordingly. An accused state, therefore, will face some reputational loss, but the lower the estimated probability that a violation has taken place, the smaller the impact on its reputation will be. Not only will violating states face a reduced sanction as a result but also some nonviolating states that are suspected of having committed a violation will suffer an undeserved reputational loss because they are wrongly suspected of having breached their international legal obligations.

The net result is to reduce the incentive to comply that the reputational sanction provides, because the consequences of violation relative to compliance are reduced by the uncertainty about the actions a state takes.

#### *The Impact of Uncertainty*

So each type of uncertainty—whether it relates to the payoffs at issue, the legal rule in place, or the actions taken—reduces the incentive to comply by reducing the cost of a violation. Inferences about reputation become more difficult to draw on the basis of available information. The reputational stakes of a particular action, then, are reduced. The result is that some states that would have complied may breach because the reputational sanctions are reduced.

However, states are not helpless in the face of this reality. To the extent that they are able to promote more complete information, they help to retain the compliance pull of reputation. Among the strategies they use are efforts to clarify legal rules (e.g., codification of customary international law); transparency requirements in international agreements (e.g., the WTO's Sanitary and Phytosanitary Agreement, which governs health and safety, includes a variety of transparency requirements,

including that trade measures justified on health and safety grounds must be published prior to their enforcement<sup>20</sup> measures to ensure ongoing communication (e.g., international antitrust agreements typically provide specific information-sharing provisions, and regulators interact through the International Competition Network),<sup>21</sup> reporting and monitoring requirements in agreements (e.g., the Convention on the Elimination of Discrimination Against Women (CEDAW) requires periodic reports from states); and so on. And of course states are able to discuss conflicts and disagreements and share information directly with one another when problems arise.

This effort to improve the flow of information sounds a great deal like the sort of behavior that the "managerial school" would like to encourage (Chayes and Chayes 1995). The Chayes focus on how communication among parties to an agreement can improve compliance by resolving interpretative ambiguities; generating consensus as to what constitutes compliance; suggesting new methods of fulfilling obligations; and persuading the parties to comply with the agreement. The actual mechanism through which compliance occurs, however, is quite different in the managerial model. To the extent that model provides an account of why states comply, it asserts that states have a "general propensity to comply" with commitments. In the rational choice model developed in this book, of course, no such general propensity is assumed. The informational mechanisms discussed earlier (along with others) are useful because they increase the relevance and force of reputational sanctions. That is, better information leads to payoffs that promote compliance.

When states make informational claims, of course, the goal is not always to simply share truthful information with others. Because informational issues affect payoffs, states have an incentive to attempt to manipulate the information sets of others. This is most obviously done when a state attempts to hide its illegal conduct, but attempts to affect the views of other states are common with respect to each of the three informational categories mentioned in this section (payoffs, legal rules, and actions). For example, on March 31, 1939, Britain's prime minister, Neville Chamberlain, declared in a speech on the floor of the House of Commons that Great Britain and France would guarantee the security of Poland against a possible German invasion. This statement was intended to lend credibility to France's commitments to Poland under the 1921 Franco-Polish Military Alliance, and it foreshadowed the signing of

an Anglo-Polish mutual defense treaty in August of 1939. In effect, the speech alleged that Britain's payoffs were such that it would not stand by if Poland were attacked. The strong rhetoric of states on the subject of a customary international law of expropriation is another example in which states have attempted to influence the content of the law. For many years the United States, among others, insisted that the "Hull Rule" of prompt, adequate, and effective compensation applied when foreign investment was expropriated. Many other states resisted this notion and argued that some lower standard applied.<sup>22</sup>

So the role of information helps to explain why states expend so much effort to influence perceptions of other states and nonstate actors. These efforts include official denials (whether true or not) regarding the conduct of the state, participation in debates about existing rules of international law, and claims about the stakes involved in decisions. The more successful such efforts, the smaller the reputational consequence for a violating state.

The same considerations explain how the work of some human rights NGOs, such as Amnesty International, can affect state behavior. Credible reports that shed light on the conduct of states reduce the uncertainty regarding their compliance and, therefore, increase the reputational consequences of a violation.

Interestingly, many human rights NGOs are also engaged in an effort to push the frontier of human rights law and expand the legal rules to include more types of conduct. To this end, they advance arguments about why particular actions should be considered human rights violations, or they may even assert that a particular rule exists even when a disinterested observer might disagree. This effort is intended to generate greater consensus on human rights obligations and persuade at least some people that additional human rights norms should be considered legal obligations. This behavior may not generate any increase in certainty (if successful, it may simply expand the set of actions that is considered illegal without generating any additional clarity at the edges), but it can serve to expand the set of behaviors that international law proscribes.

One result of these actions is that the difference between a violation of a human rights treaty—where the law is relatively clear—and the violation of an alleged rule of customary law, where the rule is often less clear, is not emphasized. This generates a tension for some human rights groups. When reporting on human rights violations, it is important that

they retain a high level of credibility. If a group is perceived to be highly partisan or overly eager to condemn the practices of governments, its reports will lose credibility and cease to be useful tools of information dissemination. When attempting to extend the boundaries of the law, however, a human rights group has an incentive to argue for an expansive definition of human rights. Though the groups need some credibility here as well, this work is more in the style of advocacy and less in the style of disinterested reporting. A group that engages in its advocacy function in an evenhanded and neutral fashion will have less success influencing the course of human rights law. But a group that pushes too hard in its advocacy work will be less credible when it attempts to shed light on existing violations.

### The Compartmentalizing of Reputation

To say that states want and pursue a reputation for compliance with international law leaves open the question of whether a state has a single reputation for compliance or whether that reputation varies by subject area, by who the counter-party to an agreement is, or by regime. Some commentators argue that states have different reputations with respect to different agreements (Downs and Jones 2002; Mercer 1996, p. 7), while others speak of states having only a single reputation (Chayes and Chayes 1995).<sup>23</sup> Most of the following discussion concentrates on the question of whether states have different reputations in different issue areas. Analogous discussions could investigate the extent to which different regimes within a state have different reputations, and to which different regimes within a state have different reputations, and the extent to which reputation depends not only on the acting state but also on the observing state. Because the same basic ideas apply to these other dimensions along which compartmentalization of reputation may take place, I discuss them in a more abbreviated way after considering issue-specific reputations.

#### *Reputation by Issue Area*

The extent to which violations affect a state's reputation generally, and the extent to which the impact is limited to one issue area, depends primarily on what it is that observing states learn as a result of a violation. When a state violates a commitment, it reveals that its expected

payoff from a violation is larger than its expected payoff from compliance. Observing states must try to infer why this is so. One possibility is that the direct gains from violations relative to compliance are relatively large. If this is so, there is no reason for a violation in this issue area to affect other issue areas. Suppose, for example, Venezuela enters into and subsequently violates an environmental agreement. If observing states conclude that this violation reflects a Venezuelan view that future environmental harms will be small (and therefore Venezuela is not prepared to bear significant costs to protect the environment), then those observing states will update their beliefs about Venezuela's attitude toward the environment. This updating will have no impact on international commitments in, say, trade or security. Another possibility, however, is that observing states will conclude that the violation reflects a more generalized reluctance to bear costs today in exchange for future benefits. It might, for example, reflect the fact that Venezuelan leaders have a higher discount rate than previously thought. If so, the violation will lead to an updating of beliefs about Venezuela's willingness to comply with international law in general and will, as a result, impact all of Venezuela's efforts to achieve international cooperation.

It is helpful to be clear about what it means to have one or several reputations. If a state has a single reputation across all issue areas, then all actions of that state impact this reputation. At the extreme, actions in the human rights area have no greater reputational impact on future human rights commitments than on, say, future economic or environmental promises. If, on the other hand, a state's reputation is fully compartmentalized by issue area, actions in one area have no bearing on the state's reputation in another area. A state can engage in widespread and flagrant violations of commitments in, for example, the security area without affecting its reputation in the economic area. Even within a particular area, a state may have a number of reputations depending on the particulars of the interaction. Huoh (1997), for example, disaggregates a reputation for resolve into four different categories.

The sensible answer to the question of whether states have one reputation or multiple reputations, given current understandings about reputation, is almost certainly "both." For several reasons, it is implausible for violations in one issue area to be strictly limited to that area. First, a violation (or compliance) provides information about a state's attitudes toward a particular area of law. It signals how important that area is to the state, how important it feels future cooperation in the area

will be, and so on. This is simple enough, but it requires some sense of what constitutes an "area" of law. Would the Geneva Convention Relative to the Treatment of Prisoners of War, for example, be its own area, meaning that violations would not implicate other areas? Or would violations of those conventions speak to a state's reputation with respect to humanitarian law generally? Would they have implications for other security issues such as a state's reputation for compliance with bans on land mines or other military technology? Or would violations instead, or perhaps also, affect a state's reputation for compliance with human rights treaties? The answer, of course, is that there is no answer. What constitutes an area for this purpose will depend heavily on context, and even with respect to a single compliance decision there will not be a single "area." A violation of a fisheries treaty may signal both that the state is relatively unconcerned about harm to the fishing stocks and that there is little domestic support for environmental measures more generally.

Second, violations of international commitments will at a minimum provide information about a state's underlying discount rate. Even if that rate varies somewhat from subject matter to subject matter, the relevant discount rates in each area will be determined in part by the state's general willingness to accept present costs for future benefits, as determined by, for example, the state's domestic structure.

Finally, observation suggests that reputations are generalized, at least to some extent. This is most clearly visible in the form of rogue states that have sacrificed whatever reputational capital they once had. These states are normally not considered to be violators in only some small number of discrete areas (though there may be a few areas that are most responsible for their loss of reputation) while simultaneously being thought of as reliable partners in other areas. It appears instead that sufficiently egregious violations in a few areas are often enough to compromise a state's reputation across the board. This is consistent with the model of reputation developed here, because egregious violations in several areas might indicate that the state has a high discount rate or simply expects few benefits from a strong reputation. If so, it is appropriate for other states to draw negative inferences about the acting state's general reputation.

If the notion that a state has multiple reputations that are completely isolated from one another seems far-fetched, it is similarly difficult to believe that states have a single reputation across issue areas.

Consider, for example, the Pinochet regime in Chile in the early 1970s. Pinochet came to power in a military coup in 1973 that removed the leftist government of Salvador Allende. The Allende government had nationalized industry, seized land and factories, and implemented price controls, all of which had combined to leave the Chilean economy in tatters and make international investors extremely anxious. Pinochet's military junta thus came to power with two goals: economic stabilization and political reform designed to prevent the recurrence of polarizing and self-destructive policies. The political aspect of the agenda was carried out through repression, much of which constituted a violation of the ICCPR, which Chile ratified in 1972 and which entered into force in March 1976. In contrast, the neoliberal economic reforms were carried out with an eye toward economic stabilization and restoration of Chile's credibility with the global financial and business communities, a program that yielded fruitful results from 1977 to 1981 and then again following Chile's graduation from IMF assistance in the late 1980s. In this example, Pinochet's Chile benefited from a reputation for compliance with economic commitments but surely had a much weaker reputation on human rights issues. The 2003 American-led invasion of Iraq offers another example. Though there was some spillover to other areas, in the main the reputational harm to the United States seems to have been limited to matters of peace and security.

So it is likely that states have different reputations in different issue areas, but that these reputations are related to one another. One can think of the ripples a stone dropped in a lake creates, with the place where the stone enters the lake being analogous to the location of a violation. The reputational consequences are greatest at the point of impact, and at points further from the specific violation, the reputational consequences diminish. Though a state is likely to have multiple reputations, it is unlikely that these reputations are entirely independent of one another. A violation of an environmental agreement, for example, may have its largest impact on a state's reputation for compliance with environmental commitments, but will often also have at least some spillover impact on other areas.

Notice that the acting state cannot control the extent to which its behavior in one area affects its reputation in other areas. The extent to which reputation spills over from one area to another is determined by the extent to which a state's actions provide information about characteristics that are relevant only to a particular context (e.g., whether the

state values ongoing cooperation on a narrow issue) as opposed to a broader set of situations (e.g., the state's discount rate). If an observing state believes that another state's breach in the human rights area signals an unwillingness to comply in the trade area, for example, then that breach will carry reputational costs in the trade area.

The question of whether states have one or several reputations has implications for the effectiveness of international agreements. Among other effects, the number of reputations a state has determines the extent to which decisions about compliance in one area are influenced by reputational concerns in other areas. Can decisions on peace and security, for example, be made without affecting present and future economic commitments? The answer to this question has important implications for the decisions governments make. Consider, for example, American efforts in the war on terror. If reputation is highly compartmentalized, the United States can be much more aggressive in some of its actions than would be the case if there were large reputational spillovers into other areas. Suppose the United States violates international rules governing the treatment of prisoners. If it suffers reputational harm only with respect to humanitarian law, the associated costs are modest. Further U.S. efforts in the humanitarian law area are unlikely to be of central importance to the country, so even if reaching agreement in this area is more difficult in the future, the costs to the United States are very small. The important caveat to this point is that the United States may wish to use humanitarian law as a tool to protect its own citizens and soldiers and current U.S. violations may hamper this effort. If, on the other hand, U.S. violations produce reputational spillovers that affect other areas of international cooperation, the costs might be higher. If, for example, U.S. violations hinder negotiations regarding North Korea's nuclear ambitions, the costs are much higher.

One area where the compartmentalizing of reputation is important is human rights. Consider, for example, the potential for international law to constrain the human rights conduct of a state. Venezuela, for example, is party to the ICCPR, which prohibits, among other things, torture, degradation, arbitrary arrest, and detention. It also guarantees rights to certain legal processes. The breakdown in recent years of law and order in Venezuela has compromised these rights on a broad scale, and prompting inquiries by the UN Commission on Human Rights as well as the Inter-American Commission on Human Rights. If reputation is fully compartmentalized, these developments impose reputational

consequences only on Venezuela's human rights commitments. Unless those commitments are linked to international benefits that the leaders of Venezuela care about (e.g., international trade), it is hard to see why that reputational loss would concern them. If, on the other hand, the state suffered a general loss of reputation, this would be costly to decision-makers, who would find it more difficult to pursue whatever other objectives they may have in the international arena.

Generalizing from these examples, if reputations are compartmentalized, it will be more difficult to generate compliance in areas where states have little interest in building a good reputation. It will, conversely, be easier to do so in areas where states have much to gain from a good reputation. This may imply that cooperation will be difficult in, for example, human rights and easier in, for example, economic interactions.

It has been argued that the existence of a single reputation implies that compliance rates should be comparable across issue areas (Downs and Jones 2002). This, however, need not be true. Compliance rates are determined by the reputational and nonreputational costs and benefits of compliance. A single reputation does not imply that either the costs or the benefits of compliance are constant across issue areas. Consider, for example, the decision to violate an extradition treaty. Even if we assume a single reputation, the compliance decision will reflect, in addition to the state's existing reputation, the nonreputational costs and benefits and the reputational consequences of violation or compliance. There is no reason to think that the nonreputational payoffs in the extradition area bear any particular relation to those relevant to, say, a treaty governing the establishment of national borders. There may be powerful domestic pressure to ignore the extradition treaty, and this may generate nonreputational payoffs that make violation the state's best strategy. With respect to national borders, the state may have no reason to challenge the established borders, and so compliance is a better approach. With respect to reputational payoffs, the violation of an extradition treaty may be a relatively minor issue that has little impact on a state's general reputation, whereas the violation of a treaty establishing a boundary may have a profoundly negative impact on reputation. The point here is that differences in reputational consequences do not require that states have multiple reputations.

If, indeed, states have multiple reputations but they are interrelated, as discussed, what are the implications for international law and compliance? Most obviously, state actions in an area will have their greatest

impact within that area. So a decision to violate a trade agreement will, first and foremost, affect a state's reputation in trade and especially with respect to the agreement at issue. But the reputational consequences will not be limited to one narrow area. There will be spillover to other "adjacent" issue areas such as other economic agreements. For more distant issues such as environment or human rights, the reputational impact will presumably be still less. So while a state's reputations may differ from one area to another, they are not entirely independent of one another.

When making a compliance decision, then, a state will take into account the reputational impact of its actions across all issue areas, but will also recognize that the reputational sanctions will be largest in the areas closest to the one at issue.

#### *Reputation by Regime*

Once we recognize that states are likely to have different if interdependent reputations across issue areas, it is easy to see that reputations may vary along other dimensions as well. For example, the reputation of a state is likely affected by the regime in place. This is most obvious in the case of extreme regime changes. To illustrate, consider again the example of Chile during the transition from the Allende to the Pinochet government. Under Allende, Chile expropriated the assets of foreign firms and, accordingly, had a reputation among both states and potential investors for disregarding the existing international law on investment.<sup>24</sup> When Pinochet seized power, this reputation changed almost instantly, and foreign investment returned.

When regime changes are as dramatic as in this example, it is easy to see why a state's reputation would change. Though less dramatic, it is likely that more modest regime shifts also affect a state's reputation. For example, the election of a Democrat to the White House might reduce the United States' reputation for compliance with its free trade agreements. Because Democrats rely on organized labor for much of their political support, their commitment to free trade agreements such as the recently signed DR-CAFTA might be in doubt. To be sure, shifts of this sort occur only at the margin, and the new regime retains an interest in honoring the commitments made by its predecessor regimes. The point here is simply that some reputational effects can come about in this way.

#### *Reputation by Dyad*

Just as reputation may vary from one issue area to another, it is conceivable that a country's reputation (or its reputation in a particular area) varies from one counter-party to another. There is, after all, no reason to think that a state will behave in the same way toward its friends and allies as it will toward its enemies and rivals. If a state's willingness to comply varies based on these factors, rational states will take this into account when they deal with one another. In negotiations with Canada, for example, the United States benefits from a high level of reputation and trust. In dealings with Iran, the United States almost surely enjoys a less positive reputation. On the basis of a review of case studies on the subject of security and deterrence, Huth (1997) concludes that there is some evidence to support the hypothesis that reputations form within country-pairs. The extent to which this evidence speaks to reputation in international law is, however, an open question.

In the context of my model of reputation, it is possible to understand dyad-specific reputations as a reflection of the fact that states have more frequent and more valuable interactions with some counter-parties than with others. As mentioned, the United States and Canada have frequent interactions, many of which yield high payoffs from cooperation. The United States has fewer interactions with, say, Tanzania. In terms of the earlier discussion of the value of reputation, the value the United States places on a reputation may be different when it deals with Canada than when it deals with Tanzania. The United States, then, would have one value function (which could be labeled  $V_C$ ) with respect to Canada and another with respect to Tanzania (which could be called  $V_T$ ). Like compartmentalized reputations by issue area, there is no compelling theoretical reason to believe that a state will have either a single reputation across all counter-parties or an entirely compartmentalized one. Both of these outcomes, and many that lie between these extremes, could generate a stable equilibrium.

That said, it seems unlikely that a state's reputation is entirely compartmentalized by counter-party. At a minimum, by observing the behavior of a country with its other partners, a state can draw lessons about that country's discount rate and its attitude toward its legal obligations. Furthermore, if reputation were entirely dyad specific, we would expect more default in instances where one state places little value on its future ability to make credible commitments with its

counter-party. For example, when a developing country accepts a World Bank loan, its promise to repay is backed by the country's reputation (at least its reputation with respect to financial obligations). Though it is certainly within the legal capacity of the state to default on the loan, countries have shown themselves reluctant to take such a step for fear that this might generate a negative signal to other potential lenders, including both private lenders and other states. Notice that this is not simply a concern about signaling that the state is in poor financial health. Where default on World Bank or IMF loans is being considered, it is often well known that the state is in dire financial straits. If anything, from a purely financial perspective, defaulting on existing debts would make a state more able to repay because the default would reduce its total debt load. Nevertheless, states do everything they can to avoid defaulting, because they fear the reaction of international financial markets, including parties other than the World Bank or IMF. Their reputational concern certainly extends beyond the specific contractual parties involved. Third parties observing the default will draw inferences about the likelihood of future defaults from the state's conduct (Tomz 2007).

Trade offers another example. When entering into negotiations regarding DR-CAFTA, the Central American participants would be foolish to ignore the behavior of the United States under the WTO and NAFTA. These other agreements suggest, for example, that the United States generally (with some exceptions) respects the decisions of arbitral bodies established by trade agreements.<sup>25</sup>

A more dramatic example is the birth of a new state. For example, after the collapse of the Soviet Union, Ukraine became an independent state. It would have been foolish for Ukraine to treat all other states as having a clean reputational slate. Ukrainian officials had knowledge of how states had reacted to legal obligations in the past, even if none of those obligations had been with Ukraine itself. By taking into account the past actions of states, Ukraine was able to develop a better estimate of the reliability of its future partners.

#### *The Impact of Multiple Reputations*

In its most disaggregated form, then, it is possible that every regime of a country has its own set of reputations and that these reputations are different in each issue area and with each partner. And one could

probably identify further dimensions along which to compartmentalize reputation. As argued, however, a rational state forms its expectations about the conduct of another state on the basis of all available information, so there is no reason to think that each of these reputations is fully independent of the others. Thus, even if there is a meaningful difference in, for example, Thailand's reputation on economic commitments and its reputation on human rights commitments, and even if each successive regime in Thailand has a set of reputations that are to a certain extent distinct from that of previous regimes, there will still be spillovers: from one reputation to another. The most useful information for an observing state will obviously be the past behavior of the current regime within the relevant issue area. But what would an observing state do if there was no such past conduct? A new regime may have not yet addressed, say, human rights issues. It seems clear that absent unusual circumstances, an observing state would take into account the reputation of the prior regime on human rights issues. It would then adjust that reputation to reflect what it knew about the current regime. If the current regimes accepted and honored international obligations in other areas, even when breach offered a payoff, then the observing state might adjust its reputational estimate upward. In this simple example, both the prior regime and other issue areas have influenced perceptions.

To the extent that reputations can be isolated, however, it is interesting to ask if the force of reputation is reduced if there are few spillovers to other issue areas (or other partners, or other regimes). Downs and Jones (2002) argue that the presence of multiple regimes implies that reputation has less force.<sup>26</sup> This conclusion, however, requires an assumption that as a state's reputation narrows, its force within the area in which it is relevant does not increase. Consider the following example. Japan and Russia have both trade and environmental dealings with one another. Assume that each state has a single reputation and that Russia has failed to honor an environmental agreement with Japan. Its reputation will be hurt as a result, and this reputational sanction will impose costs on Russia when it deals with Japan on either environmental or trade issues. For clarity, assume that the cost of this lost reputation is  $E$  in the environmental context and  $T$  in the trade context, leading to a total cost of  $E + T$ .

Now assume that the states have separate reputations for trade agreements and environmental agreements, and that a violation of an environmental agreement has no impact on a state's reputation in the

trade area. Assume again that Russia has violated an environmental agreement. By assumption, there is no reputational consequence in the trade arena (i.e.,  $T = 0$ ). There remains, of course, a reputational consequence in the environmental area. If we assume that the cost in the environmental area is the same in this case as it would be if the state has a single reputation, then the total cost of the violation is simply  $E$ . It follows that the assumption of multiple reputations implies that reputation has less impact on compliance.

But this analysis is flawed. The reputational cost in the environmental area will be higher if there are multiple reputations than if there is a single reputation. This is so because if there are multiple reputations, there are necessarily fewer "compliance opportunities" for a state to establish a reputation in a given area. If Russia has a single reputation, Russian actions in the trade area yield information that allows Japan to form beliefs about Russian behavior in the environmental area. This means that Japan observes a larger number of compliance decisions by Russia and, therefore, its view of Russia's reputation is more stable. In contrast, if Russia has a reputation for compliance with environmental agreements that is independent from all other areas, Japan will have observed many fewer compliance decisions by Russia. Therefore, when Russia does make such a decision, the priors of observing states (including Japan) will be weaker, and the impact of a particular conduct on reputation will be larger. The Russian violation of the environmental agreement, then, will generate a reputational cost in the environmental area of some amount,  $E^* > E$ . Without additional assumptions, there is no way to know if  $E^*$  is larger or smaller than  $E + T$ .

So we cannot conclude that having multiple reputations somehow reduces states' incentive to comply with their obligations. We can conclude that reputation will have a larger impact in areas where the value of cooperation is high. In these areas, a loss of reputation imposes relatively large costs, so breach is discouraged. This might describe, for example, trade. Though there are plenty of exceptions, it is clear that states tend to comply with their trade obligations even when these obligations impose short-term costs. It seems likely that consistently breaching trade obligations frustrates a good deal of future cooperation and, therefore, is costly. Human rights may represent a useful example in which the opposite is true. States that are tempted to violate human rights commitments are unlikely to be concerned about the possibility that future human rights agreements will be difficult to negotiate or join. This

reduces the costs associated with the violation of these treaties and, therefore, makes such a violation more likely.

#### Limits and Caveats

Before discussing the sources of international law in the next two chapters, it is useful to address some of the limits and questions raised by the theory advanced up to this point. The following subsections discuss the limits on the ability of reputation to influence state conduct and the relationship between a reputation for compliance with international law and other kinds of reputation.

#### *The Limits on Reputation's Ability to Generate Compliance*

It should be clear from the foregoing discussion that reputational sanctions can generate compliance in some but not all circumstances. This is a necessary feature of any model of international law. We observe that states sometimes violate their obligations, and so our theory must have some explanation for when that will happen.

Consider one of the more dramatic failures of international law in the last century—the Munich Accord of 1938. Under the terms of this agreement, France and Britain agreed to the German annexation of the Sudetenland, a German-speaking region of Czechoslovakia. In exchange, Hitler promised that he had no further territorial ambitions in Europe. The Munich Accord failed to deter Germany's subsequent decision to seize the rest of Czechoslovakia, or the later decision to attack Poland, which triggered the outbreak of World War II. Nazi Germany's failure to comply with the Munich Accord was, to be sure, a failure of international law, but it is not one that should surprise us. It is clear (at least in retrospect) that Nazi Germany had no reason to value a good reputation. Hitler's ambitions required that he ignore international legal norms, including the national boundaries of other states. These ambitions were inconsistent with the maintenance of a good reputation for compliance with international law, so there was no point in cultivating such a reputation. In other words, the harm to Germany's reputation as a result of its violation of the Munich Accord imposed only a modest cost on the state.

Probably more important than Nazi Germany's lack of interest in a strong reputation was the fact that the decision to violate the Munich

Accord involved calculations about national security, territorial expansion, and European dominance. Though states have a reason to protect their reputation for compliance with international law, the value of reputation is, nevertheless, limited. Even the threat of a total loss of reputation would not normally cause a state to take actions that threaten its very existence. In the lead-up to World War II, Germany sought expansion and control of much of Europe, a prize whose value surely exceeded the reputational consequences of Hitler's actions. The pre-World War II experience in Europe also illustrates how observing states update their beliefs. As Nazi Germany repeatedly failed to honor commitments, other European states adjusted their estimates of Germany's reputation, eventually concluding that they could neither seek nor rely on further German promises.

Indeed, while the violation of the Munich Accord and the seizure of the remainder of Czechoslovakia may have been the signal to Europe that Germany did not intend to honor its own agreements or the territorial integrity of other European states, there had been earlier warning signs. In 1938, prior to Munich, Germany had violated one of the terms of the Treaty of Versailles when it annexed Austria.<sup>27</sup> Earlier violations of the Treaty of Versailles included Germany's remilitarization of the Rhineland in 1936, and Germany's rearmament throughout the mid-1930s. Nevertheless, many western European nations had come to see the Treaty of Versailles as unnecessarily harsh and punitive, and had higher expectations of Germany's compliance with the Munich Accord. So it was the violation of the latter agreement that caused the collapse of Germany's reputation and prompted the British and French to issue the March 1939 guarantee of Polish security against German aggression.

Generalizing this example yields two important lessons about reputation and international law. First, the force of reputation is affected by a state's interest in developing or maintaining a good reputation. States that anticipate little or no return from investments in reputation are less likely to comply with international law. The effectiveness of reputation hinges on the state's desire to be able to make credible promises in the future.

Second, the force of reputation is limited. Even a state eager to nurture its reputation will violate its legal commitments if the nonreputational payoff from doing so is large enough. This implies that when the stakes are very high, the likelihood that reputation can influence the outcome is smaller. Certainly when states have honest concerns about

fundamental security interest, for example, international law is unlikely to have much influence on their decisions. If this is correct, we should expect to see (as we do) international law largely put aside during moments of extreme national crisis. Something similar is true of domestic law, where it is sometimes said that *inter arma silent leges* ("in times of war the laws are silent"). In such moments of crisis, more powerful forces than international law are likely to push it aside. As the stakes at issue grow smaller, however, the likelihood that reputational issues might tip the balance toward compliance grows.

The weakness of international law in the face of high-stakes decisions does not imply that international law cannot be usefully deployed in high-stakes areas, however. It seems improbable, for example, that international law can prevent a war that would otherwise take place, but that does not mean that international law is irrelevant to all issues related to warfare. To take one example, though it is unlikely to prevent a war, international law may be able to improve the treatment of prisoners of war. The Geneva Conventions regulate the treatment of prisoners of war and, though they are certainly not always complied with, impose reputational costs on states that choose to ignore the conventions. Provisions of the Geneva Convention Relative to the Treatment of Prisoners of War stipulate that a detaining power must allow visits to prisoners of war by representatives of "protecting powers," that these visitors must be allowed to interview prisoners, and must be allowed to visit the premises in which they are housed.<sup>28</sup> These requirements make it possible for violations of the Geneva Convention to become public and provide the detaining power with an incentive to comply. Notice that even the act of denying access to visitors is a violation that comes with an attendant reputational cost. For example, following revelations of detainee abuse at the Abu Ghraib prison in Iraq, the U.S. army suggested that the Red Cross should no longer be permitted to make no-notice visits to cellblocks. This suggestion became public, and was a source of much embarrassment to the United States.<sup>29</sup> Where this form of reputational cost is insufficient to generate compliance, reciprocity may succeed. Both sides in a conflict may prefer a regime of mutual access to prisoners to one of no access. The international law relating to prisoners of war, then, has some chance of affecting state behavior, because the gains from a violation may be outweighed by its reputational costs (Morrow 2001).

More generally, law can matter in high-stakes areas by targeting individual decisions that do not themselves implicate high stakes for the

parties. Compliance with rules governing the treatment of prisoners of war will often (though not always) represent a modest cost, and legal obligations might influence the cost-benefit calculation. The lesson is that international law cannot easily influence high-stakes decisions. When trying to create binding obligations, then, negotiators are well advised to find ways to constrain states by influencing decisions with lower stakes. An example is the use of inspections in arms control regimes such as the NPT. A variety of protocols (the International Atomic Energy Agency Protocols), signed pursuant to the NPT, commit states to certain levels of transparency regarding their use of nuclear technology. The stakes involved in complying with these protocols individually are much lower than the relatively high stakes involved in developing a nuclear capability. Thus, reputational sanctions associated with the protocols may deter violation of a protocol, and thus of the larger arms control agreement, when such reputational costs alone would not have prevented a state from defecting from the arms control regime as a whole. If high-stakes decisions are difficult to influence, one might hope that low-stakes decisions are easy to affect. That is, one might think that a decision to violate the law imposes some fixed reputational sanction and, therefore, that compliance decisions involving modest costs and benefits are systematically more likely to be affected by international law than are decisions that implicate larger stakes. This is only true, however, if one makes the implausible assumption that reputational sanctions are unrelated to the stakes in the case.

It seems much more realistic to assume that a violation of an obligation that is of fundamental interest to other states will generate a larger reputational sanction than will a violation of a minor obligation. The Kyoto Protocol, for example, requires that state parties submit an annual inventory of greenhouse gas emissions as well as greenhouse gas removal by sinks. A state that fails to meet its reporting requirements on time and instead submits the required reports a day late is unlikely to face reputational sanctions of any consequence. In contrast, the same agreement requires a certain percentage reduction in the emission of greenhouse gases, with different countries having different targets. A decision to ignore this obligation would result in reputational sanctions. To put it more generally, some violations of international law will simply be too trivial to trigger reputational sanctions, even when other states can observe the violation. In these cases, reputation will have little or no effect on a state's decision to comply. It is in cases in which

observing states view the obligation as a significant basis for forming future expectations of compliance, and in which the stakes for the state deciding whether to comply are sufficiently low, that reputation will be able to exert the most influence.

#### *Other Kinds of Reputation*

The discussion of reputation up to this point has been focused on a state's reputation for compliance with international law, which could be called (somewhat awkwardly) the state's "compliance reputation." The basic argument relies on the notion that states wish to be able to make and elicit credible commitments and are willing to forgo at least some short-term gains in order to develop a reputation that allows them to make such commitments.

This focus on a state's reputation for compliance with international law might leave the impression that a state has no other reputational concerns. Indeed, one criticism that has been made of a reputational theory is that states and their leaders may care about many things other than a reputation for compliance with international law (Goldsmith and Posner 2009).<sup>29</sup> A state may want a reputation for toughness, retributiveness, kindness, generosity, or any number of other things (Keohane 1997). In fact, the bulk of the literature on reputation in international affairs is concerned with a state's reputation for resolve rather than for compliance with international law (Mercer 1996). Furthermore, these other reputational concerns may be in tension with whatever interest states have in respecting international law. The problem with this argument is not its premise (that states care about other things) but its conclusion. The fact that other factors are relevant tells us nothing about whether a state's desire for a reputation for compliance generates an incentive to comply with the law.

By way of example, imagine a conflict between the United States and India with respect to intellectual property. This could take the form of a dispute about India's legal obligations under the TRIPs agreement, as was the case in the *India-Patents* case at the WTO,<sup>30</sup> or it could simply be a case of the United States attempting to persuade or coerce India into a particular course of action, in the absence of a legal rule. The latter type of dispute took place during the Uruguay Round of trade talks in the late 1980s and early 1990s. The United States, among others, wanted the Uruguay Round to be approved. India was objecting in particular to

the TRIPs agreement, which would have required India to legislate increased protections for intellectual property. Both sides were concerned about the substantive issues at stake, but both sides also had reputational concerns. The United States had good reason to promote a reputation for dealing harshly with those that do not capitulate to its will, especially in the intellectual property area.<sup>31</sup> India, on the other hand, had an interest in resisting an increase in international intellectual property requirements, not least because the pre-TRIPs rules had generated many jobs for Indian citizens and opportunities for Indian businesses, and the TRIPs regime threatened these gains. Internationally, India also had an interest in building its reputation as a leader among developing states and a country that would resist coercion.

Of course, exactly the same reputational concerns were present when these two countries were involved in the already mentioned India Patents case at the WTO. That dispute was the first WTO case dealing with the TRIPs agreement, so both parties had reputational interests in resisting a compromise. After the ruling, which went in favor of the United States, India had an interest in refusing to comply, in an effort to develop a reputation as a country that is not worth pursuing on TRIPs issues. The United States had an incentive to ensure Indian compliance to demonstrate the U.S. commitment to protecting its interests under TRIPs.

None of the reputational effects discussed in these two examples speak to a state's reputation for compliance with international law. Furthermore, India's interests, as described in the preceding paragraph, are in tension with its reputational interest in compliance. This particular case ended with compliance. After the United States won before the dispute settlement bodies of the WTO, India agreed to bring itself into compliance, and did so in 1999.<sup>32</sup> The broader point is that when the compliance reputation and other reputational concerns pull in opposite directions, either of them may dominate.

This point can be made much more generally. There will frequently be many costs and benefits at play when a state is making a decision about compliance. Sometimes the state will have sufficient incentives to comply without worrying about its compliance reputation. Other times the incentive to violate a commitment will be sufficiently large that the state will do so despite its concerns about its compliance reputation. Finally, there will be times when a state's reputational interests with respect to compliance will cause it to comply despite the fact that the

other relevant costs and benefits provide an incentive to violate the commitment.

What all of this really means is that there are many influences on a state's behavior beyond its compliance reputation. This observation does not, however, undermine the basic point that the compliance reputation is relevant to state decisions. Like all influences, compliance reputation operates at the margin—putting a thumb on the scale in favor of compliance. If other incentives to violate a norm are sufficiently strong, reputation will not prevent violation. It matters most when the decision to violate or comply is a relatively close call. In these cases, reputation can generate compliance where there might otherwise have been violation.