

REPUTATION, COMPLIANCE, AND INTERNATIONAL LAW

GEORGE W. DOWNS and MICHAEL A. JONES*

ABSTRACT

Increasingly skeptical about the efficiency and effectiveness of formal multilateral enforcement mechanisms, a growing number of international relations theorists and international lawyers have begun to argue that states' reputational concerns are actually the principal mechanism for maintaining a high level of treaty compliance. This essay argues that there are a number of empirical and theoretical reasons for believing that the actual effects of reputation are both weaker and more complicated than the standard view of reputation suggests. While states have reason to revise their estimates of a state's reputation following a defection or pattern of defections, they have reason to do so only in connection with those agreements that they believe are (1) affected by the same or similar sources of fluctuating compliance costs and (2) valued the same or less by the defecting state. Among the implications of this is that all but the newest states maintain multiple reputations.

I. INTRODUCTION

INTERNATIONAL relations theorists and international lawyers have long argued that reputational concerns help ensure that states maintain their agreements. In the past decade, a growing preoccupation with international cooperation and its analytical underpinnings have combined to increase the theoretical centrality of reputation. In the wake of works such as those by Robert Keohane, Robert Axelrod, and Paul Milgrom, Douglass North, and Barry Weingast, it now stands as the linchpin of the dominant neoliberal institutionalist theory of decentralized cooperation.¹ According to the stan-

* Downs is Professor, Department of Politics, New York University. Jones is Assistant Professor, Department of Mathematics, Montclair State University, and Visiting Scholar, Department of Politics, New York University. The authors wish to thank David Rocke, Robert Keohane, Michael Gilligan, Bruce Bueno de Mesquita, Eric Posner, Jack Goldsmith, Alan Sykes, an anonymous referee, and participants in the conference Rational Choice and International Law, University of Chicago Law School, April 2001, for their comments on this essay and related papers.

¹ See Robert O. Keohane, *After Hegemony: Cooperation and Discord in the World Political Economy* 105–8 (1984); Robert Axelrod, *The Evolution of Cooperation* (1984); Paul R. Milgrom, Douglass C. North, & Barry R. Weingast, *The Role of Institutions in the Revival of Trade: The Medieval Law Merchant, Private Judges, and Champagne Fairs*, 2 *Econ. & Pol.* 1 (1990).

[*Journal of Legal Studies*, vol. XXXI (January 2002)]

© 2002 by The University of Chicago. All rights reserved. 0047-2530/2002/3101-0013\$01.50

dard argument, a major—if not the major—reason why states keep commitments, even those that produce a lower level of returns than expected, is because they fear that any evidence of unreliability will damage their current cooperative relationships and lead other states to reduce their willingness to enter into future agreements.² Since the opportunity costs associated with this forgone cooperation are substantial, the vast majority of states possess a strong incentive to behave cooperatively.³

There is, however, empirical evidence to suggest that the impact of reputation is either weaker or more complicated than much of the theoretical literature suggests. While compliance rates are relatively good in general, they are often considerably lower than one would expect them to be if every defection had important implications for every current and future agreement.⁴ This is true even if we allow for the effects of imperfect information. Even more suspicious is the fact that it is common for a given state to evidence very different levels of compliance reliability—the building block of reputational inference—in connection with different agreements. It is difficult to believe that it is efficient for other states to simply ignore this fact and treat each state as if it possessed only a single reputation.

Certainly, in everyday life, differentiated historical experience often results

² International legal theorists use reputation to refer to both (1) the extent to which a state is considered to be an honorable member of the international community and (2) the degree to which a state reliably upholds its international commitments. This tends to create confusion because both can, at least in theory, affect a state's compliance rate, and the first is to some extent a function of the second. We say "to some extent" because no one has yet produced a theory that explains how the two are linked. The focus of this paper is the second, rational choice meaning of the term, which should be important regardless of whether or not states are sometimes motivated to comply by a fear of international opprobrium. The reliability with which a state abides by its commitments determines its value as a prospective partner and what others are willing to commit in return. It is this rational dimension of reputation that is chiefly of interest to economists and most political scientists.

³ See Keohane, *supra* note 1. Reputation operates outside the boundaries of formal treaty law as well. It plays a critical role in determining the effectiveness of both tacit bargaining and diplomatic exchanges, or what is generally termed "cheap talk" in the literature. See Anne E. Sartori, *The Might of the Pen: A Reputational Theory of Communication in International Disputes*, 56 *Int'l Org.* 121 (2002). Reputation expectations are also central to William English's explanation of why American states agreed to repay their foreign creditors even after these creditors were prevented from seeking redress in federal courts during the 1840s. See William B. English, *Understanding the Costs of Sovereign Default: American State Debts in the 1840's*, 86 *Am. Econ. Rev.* 259 (1996).

⁴ The literature on the compliance rates associated with different areas is considerable. See Robert E. Hudec, *Enforcing International Trade Law: The Evolution of the Modern GATT Legal System* (1993); Thomas O. Bayard & Kimberly Ann Elliott, *Reciprocity and Retaliation in U.S. Trade Policy* (1994); Beth A. Simmons, *International Law and State Behavior: Commitment and Compliance in International Monetary Affairs*, 94 *Am. Pol. Sci. Rev.* 819 (2000); Edwin M. Smith, *Understanding Dynamic Obligations: Arms Control Agreements*, 64 *S. Cal. L. Rev.* 1549 (1991); Randolph M. Siverson & Joel King, *Attributes of National Alliance Membership and War Participation, 1815–1965*, 24 *Am. J. Pol. Sci.* 1 (1980); George W. Downs, *Enforcement and the Evolution of Cooperation*, 19 *Mich. J. Int'l L.* 319 (1998); *The Implementation and Effectiveness of International Environmental Commitments: Theory and Practice* (David G. Victor, Kal Raustiala, & Eugene B. Skolnikoff eds. 1998).

in individuals and organizations being assigned what are effectively multiple reputations that operate to limit the reputational consequences of a given incident. You have hired a teenager to construct a Web site. Although the teenager claims to be an expert HTML programmer, he has been foiled by rudimentary problems. If the teenager is not someone you know very well and who has only just moved into the neighborhood, you might infer that he is not trustworthy and assign him that reputation. As a consequence, you might very well dismiss an idea you have been discussing with him earlier that day whereby you would pay him to take care of your house, car, and pets when you were in Greece for the months of January and February. If, however, the teenager is someone whom you have known since he was a child and who has faithfully carried out work that you have paid him to do in the past, it is more likely that you will draw the much narrower reputational inference that he simply knows much less about programming than he thinks he does. As a consequence, you make a mental note to be extremely cautious with regard to allowing the teenager to tutor your young daughter in connection with her introductory programming class. However, you do not feel the need to set aside the caretaking agreement that you two had previously discussed.

This paper argues that the reputational consequences of a state's noncompliance with a given treaty are similarly limited by the history of its cooperative relationships with the other member states. While states have reason to revise their estimate of a state's reputation following a defection or pattern of defections, they have reason to do so only in connection with agreements that they believe (1) are affected by the same or similar sources of fluctuating compliance costs (or benefits) and (2) are valued the same or less by the defecting state.⁵ Over time, states develop a number of reputations, often quite different, in connection with different regimes and even with different treaties within the same regime.

The surplus value produced by a relationship is important because it offsets the invariably fluctuating costs of compliance. If a relationship is of little value to a state and produces only a small surplus, the slightest increase in compliance costs will lead to defection. A relationship that yields a higher surplus will not be disrupted by such a minor fluctuation in costs because compliance still produces a positive net benefit. It follows that, in general, states (1) will defect from low-net-benefit relationships more frequently than they will from high-benefit relationships and (2) will defect from low-benefit relationships whenever they defect from high-benefit relationships but not vice versa.

This implies that in treaties where the benefits of cooperation are excludable, reputation tends to protect powerful states the most, since large states

⁵ While both costs and benefits of an agreement can vary stochastically, we will focus on costs in what follows to keep things as simple as possible.

are most likely to lie at the center of the most valued cooperative relationships. In multilateral public goods agreements such as those that are prevalent in the area of environmental regulation or human rights, the reputational consequences of defection depend on the size of the treaty and the relative importance that states assign to it.

The implications of this theory for the power of reputation to ensure compliance with international law generally are substantial. Chief among these is the fact that, except in the case of new states and regimes that are believed to constitute sharp departures from the historical norm, the reputational implications of noncompliance are narrower and hence smaller than much of the literature suggests. This means that even in an increasingly integrated international system, reputational concerns cannot by themselves begin to ensure a high level of compliance with every international agreement. At the same time, however, reputational concerns are an important force for compliance in connection with certain agreements.

II. THE TRADITIONAL APPROACH TO REPUTATION

As in game theory generally, a player's reputation is viewed here as a summary of its opponents' current beliefs about the player's compliance strategy or set of strategies in connection with various commitments.⁶ The higher its opponents' expectation about how reliably a state will maintain its commitments, the better the state's reputation. Joel Sobel is among the first to explicitly connect reputation and reliability within the context of cooperation.⁷ Most earlier papers on reputation equated reputation with the ability of an incumbent monopolist to deter challenges by smaller firms; these earlier papers are part of the "chain-store" tradition, referring to the chain-store paradox of Reinhard Selten.⁸

Although reciprocal and reputational responses to defection are sometimes lumped together in the literature, we will distinguish between them in what follows. The former category refers to retaliation by the "victim" state(s) immediately following a defection and/or the punishment meted out by the international organization charged with managing a given agreement. The latter refers to the negative consequences that follow from states readjusting their estimates of the defecting state's expected reliability. This is in keeping with the common language definition of reputation that focuses primarily on a future beyond the immediate, on agreements other than the one in which defection has taken place, and on the response of a community of actors

⁶ Drew Fudenberg & Jean Tirole, *Game Theory*, ch. 9 (1991).

⁷ Joel Sobel, *A Theory of Credibility*, 52 *Rev. Econ. Stud.* 557 (1985).

⁸ Reinhard Selten, *The Chain Store Paradox*, 9 *Theory & Decision* 127 (1978).

who are either potential cooperative partners of the defecting state or partners in a different agreement that might potentially be affected.⁹

Keohane was one of the first to distinguish between the reciprocal expectations each possessed about the other's short-term response to a violation and expectations about the broader impact that a violation would have via the response of states outside the agreement. "Regimes rely not only on decentralized enforcement through retaliation but on government desires to maintain their reputations."¹⁰ "For reasons of reputation, as well as fear of retaliation and concerns about the effects of precedents, egoistic governments may follow the rules and principles of international regimes even when myopic self-interest counsels them not to."¹¹ In the economics literature, Michihiro Kandori also takes care to distinguish between personal enforcement or the immediate retaliation by the victim and sanctioning behavior of the wider community in response to information about defection.¹²

Having settled at least provisionally on a working definition of what constitutes a reputational effect, we can begin to examine the central question for both international lawyers and international relations theorists alike: How much more cooperation exists than would otherwise be the case because states must worry about their reputation? While positions vary somewhat, the dominant view in the literature is that reputation plays an extremely important role in promoting compliance.¹³ According to the traditional theory, states carry a general reputation for cooperativeness that determines their attractiveness as a treaty partner both now and in the future. Because a defection in connection with any agreement will impose reputation costs that affect all current and future agreements, states are motivated to comply with their commitments even in circumstances where they would otherwise defect.

Thus, the literature is rife with references to the cost a state will pay for losing its reputation for reliably maintaining its commitments and being "law abiding." Reflecting on what one should make of the exemplary record of

⁹ Another related reason for treating reputation and reciprocal retaliation separately is to avoid inference problems that can lead to faulty prescriptions. For example, empirical researchers in the institutionalist tradition regularly conclude that reputation is responsible for whatever decentralized cooperation they discover. This suggests that cooperation will be best increased by adopting policies that will increase system transparency on the grounds that anything that increases the level of community awareness about compliance will operate to increase the salutary effects of reputation. Yet if this "reputational" effect is due to a reciprocal response—or expectations about a reciprocal response—such a policy will be ineffective.

¹⁰ Keohane, *supra* note 1, at 108.

¹¹ *Id.* at 106.

¹² Michihiro Kandori, *Social Norms and Community Enforcement*, 59 *Rev. Econ. Stud.* 63 (1992).

¹³ Authors often oscillate in their characterization of reputational effects between what we will call the traditional view that states possess a single reputation and a more regime-specific perspective in which states have a different reputation in connection with different regimes. For this reason, our association of a given author with a given perspective should be taken only as indicating our interpretation of a specific text.

compliance associated with most agreements, Scott Barrett states, “A damaged reputation resulting from noncompliance can make it difficult for a deviant to enter into future agreements. Even a single deviation carries the risk of precipitating general erosion in law abidance, to the detriment of all states.”¹⁴ In speaking about the implications of India’s failing to adhere to the Comprehensive Test Ban after its prominent multidecade campaign for its passage, Richard Williamson speaks of the “severe costs to the nation’s reputation and perception as a trustworthy member of the international community.”¹⁵ Andrew Hurrell and Benedict Kingsbury argue that “states generally comply with international obligations . . . because of their broader concern with their reputation as reliable partners and their interest in a rule governed . . . international system.”¹⁶ And Richard Parker speculates that states complied with the International Convention for the Conservation of Atlantic Tuna Tuna-Dolphin Program even after suspending their formal participation in order to “preserve their reputation as reliable negotiating partners.”¹⁷

Abram and Antonia Chayes also believe that the power of reputational concerns to promote compliance is considerable and rivals the deterrent effect of expectations about reciprocal defection: “But in international organizations, as in other political settings, specific reciprocity is not the only or even the most important form of exchange. When a member of an organization goes back on a commitment, it compromises in some degree its reputation as a reliable partner and jeopardizes its ability to continue to reap organizational benefits.”¹⁸ Elsewhere, in support of the power of reputation, they reference Oran Young’s argument that the forces of “social opprobrium” and “the sense of shame or social disgrace” will work to induce treaty compliance.¹⁹

No detailed justification of the traditional theory of reputation exists in the literature, so one can only speculate about the rationale for thinking that states possess a single, unitary reputation for cooperation. This is the assumption that states possess a single reputation for cooperation that characterizes its expected reliability in connection with every agreement to which it is a party.

¹⁴ Scott Barrett, *International Cooperation and the International Commons*, 10 *Duke Envtl. L. & Pol’y F.* 131, 139 (1999).

¹⁵ Richard L. Williamson, Jr., *Law and the H-Bomb: Strengthening the Nonproliferation Regime to Impede Advanced Proliferation*, 28 *Cornell Int’l L. J.* 71, 72 (1995).

¹⁶ Andrew Hurrell & Benedict Kingsbury, *The International Politics of the Environment: An Introduction*, in *The International Politics of the Environment 1* (Andrew Hurrell & Benedict Kingsbury eds. 1992).

¹⁷ Richard W. Parker, *The Use and Abuse of Trade Leverage to Protect the Global Commons: What We Can Learn from the Tuna-Dolphin Conflict*, 12 *Geo. Int’l Envtl. L. Rev.* 1, 2 (1999).

¹⁸ Abram Chayes & Antonia H. Chayes, *The New Sovereignty: Compliance with International Regulatory Agreements* 273 (1995).

¹⁹ *Id.* at 152.

One possible justification is that a state's general propensity to cooperate is attributable to some fundamental underlying attribute of the state such as the ideology of its leadership or its geopolitical position. The very term "general propensity to comply" seems to imply this. Under this logic, a defection anywhere has implications everywhere because it constitutes new evidence about the basic underlying nature of the state. This dispositional vision of reputation holds some attractions for realists who tend to reify the state and are attracted to the idea of a national character, but it has several weaknesses.²⁰ One such weakness is that the idea of an immutable national character is inconsistent with the rhetoric of calculation that partisans of the unitary perspective often employ when referring to a state's realization that it needs to build a strong reputation.

A more significant weakness of this justification of the unitary-reputation theory is that it does not help us understand why states have different compliance rates in conjunction with different regimes and often with different treaties within the same regime. Instead, the character trait perspective gives us every reason to expect that a general propensity to cooperate should manifest itself in relatively uniform compliance rates, in the same way that we expect a person's general level of honesty to reflect itself uniformly across all of the dealings in which she is involved.

An alternative justification for embracing the unitary view of reputation is that it reflects a state's grand strategy. This is more consistent with the rhetoric of most unitary theorists than the reputation-as-character explanation, and it is more consistent with the contention that states generally maintain a strong reputation for compliance because it is in their interest to do so. The grand-strategy justification also appears better able than the reputation-as-character idea to explain why compliance rates vary from regime to regime or within a regime. It does not seem unreasonable to believe that a state might have some broad strategic interest in having a higher compliance rate in one area than in another or even within a regime category.

Unfortunately, the assumption that a state might be interested in maintaining different reputations in different areas contradicts the premise that states have a unitary reputation in the first place. Other states are bound to notice that the state has a higher compliance rate in some areas than in others and to condition their behavior on the state's reputation in each area and in conjunction with each treaty in the same area. If they do not do so, they will be behaving irrationally. Once this happens, the state can no longer meaningfully be said to have a single reputation in any strategic sense. It possesses multiple reputations whether it likes it or not. A state will, of course, always

²⁰ Jonathan Mercer contends that a "dispositional reputation" exists that adheres in the basic character of the actor and is independent of the actor's underlying utility function and other aspects of context; see Jonathan Mercer, *Reputation and International Politics* 6 (1996).

have an overall average compliance rate, but the other states have no reason to be preoccupied with it.

A second difficulty with the notion of reputation as a grand strategy is that it inevitably leads one to ask what strategic interest a state might have in fashioning a reputation that is less than perfect when the cost of noncompliance is believed to be so high. One possibility is that less-than-perfect reputations are the product of misperception. An imperfect technology of treaty oversight leads a treaty partner to wrongly infer that the state has defected when in reality it has not. While doubtless such mistakes occur, the likelihood that most suspected defections have really not taken place is small, in part because both sides should have an incentive to reduce any such costly mistakes over time. In any event, no one denies that most defections are real enough. Another possibility is that states are periodically defecting from agreements as part of some mixed strategy. However, this inefficient eventuality also seems unlikely since agreements are painstakingly negotiated. Some other explanation for periodic defection and imperfect levels of reliability clearly seems needed.²¹

III. THE CASE FOR MULTIPLE REPUTATIONS

One way to account for the prevalence of imperfect and variable reliability is to argue that the reputational consequences of defection are usually more bounded than the unitary theory suggests. If states have learned that the rates and timing of defections are often only weakly correlated across treaties, there is no need to readjust their reliability estimate for every treaty whenever there is a defection from a given agreement. In such a world states effectively hold not one reputation but many.

To understand why this might be the case, we first need to consider the roots of imperfect reliability in the context of cooperative relationships—a topic that has not received much attention. One explanation for this lack of attention is that international legal theorists and international relations theorists have generally been preoccupied with understanding the origins of cooperation and accounting for the high levels of compliance that generally characterize most agreements. A second reason is that for the most part both groups of theorists, to the extent that they have been exposed to formal models at all, are familiar only with models that predict perfect reliability in equilibrium.

Outside international law and mainstream international relations, however,

²¹ A number of authors implicitly abandon the idea of a unitary reputation by using the term “reputation” in a regime-specific way. However, they characteristically omit any explanation for their decision. This leaves the reader in doubt as to whether an author has a well-developed rationale for believing that a state’s defection from a treaty in one area will not have implications for its reputation in another area or whether she is simply interested in trade and has no interest in the broader question of where the boundaries of reputation are located.

the puzzle of why many cooperative agreements are punctuated by periodic defection in equilibrium has drawn some theoretical attention. Sobel, for example, describes a number of games of incomplete information where the value of cheating for player 1 varies from round to round.²² Player 2's willingness to cooperate with player 1 is a function of the latter's reliability, which can be communicated only through a history of actions. Sobel shows that if player 2 possesses less than complete information about the fluctuation of player 1's payoffs, the first player will always have a compliance rate of less than 100 percent. Player 2's uncertainty about the first player's motivations provides the first player with the opportunity to build a good reputation when the value of the stage game is low that she can later benefit by exploiting when the value of the stage game is high.²³

Sobel's insight lies in showing that variation in the utility of stage games combined with uncertainty can provide an additional explanation for less than perfect compliance. The "cashing in" feature is not essential, and most models that contain stochastic utility changes do without it. This is generally true in models of loan default, where stochastic utility changes have been used most frequently. For example, Herschel Grossman and John Van Huyck construct a game in which a state's ability to repay a loan is a function of its income, which is, in turn, partly determined by a stochastic exogenous component.²⁴ The borrower chooses to maintain a good reputation only if the expected present value of the utility it derives from validating lender expectations is at least as large as it would be if it were to repudiate its debts.

There appears to be good reason to believe that such fluctuations in payoffs are the rule rather than the exception and that they hold the key to understanding why the reliability rates associated with most agreements are less than perfect and why these rates vary from one agreement to another.²⁵ To understand the source of these fluctuations, consider the analysis of why states violate treaties advanced by the Chayeses—who seem generally to adhere to the unitary theory. They argue that the majority of treaty violations are not the product of any calculated exploitation but rather are caused by three factors: (1) the ambiguity and indeterminacy of treaty language, (2) limitations on the capacity of the parties to carry out their commitments, and (3) social, economic, and political departures from the expectation embodied in regulatory treaties.²⁶

²² Sobel, *supra* note 7.

²³ *Id.* at 564.

²⁴ Herschel Grossman & John B. Van Huyck, Sovereign Debt as a Contingent Claim: Excusable Default, Repudiation, and Reputation, 78 *Am. Econ. Rev.* 1088 (1988).

²⁵ Engaging Countries: Strengthening Compliance with International Environmental Accords (Edith Brown Weiss & Harold K. Jacobson eds. 1998).

²⁶ Chayes & Chayes, *supra* note 18, at 10.

One characteristic that these explanations have in common is that they will not necessarily have the same impact on every agreement. Treaty ambiguity and contract ambiguity are obviously contract specific in the sense that both types of ambiguity vary enormously across agreements. The capacity limitations of states to deal with agreements also vary a great deal.²⁷ The demands of the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES), for example, pose almost insurmountable detection problems for a state with as large an area and as much border traffic as the United States as compared with Switzerland. For many other agreements, the capacity demands are similar across states. Volatile economic and social consequences have a similarly diverse effect across agreements. A mild recession, for example, might have one effect on a state's propensity to fulfill its human rights commitments, another on its ability to repay its loans according to schedule, and still another on its ability to comply with a pesticide agreement.²⁸

The factors discussed by the Chayeses help us account for why a state might have different levels of reliability in connection with treaties in two different regulatory areas.²⁹ However, the same logic also holds for two different treaties in the same regulatory area. If it is known that two environmental treaties contain different amounts of ambiguity, require widely differing levels of resources to carry out, or vary in their sensitivity to different political and economic shocks, then a defection in connection with one of them need not have reputational implications for the other.

The existence of these fluctuations in compliance costs or payoffs provides states with an incentive to let their reliability rates vary, but it does not explain why treaties cannot be constructed to overcome this compliance problem. At least in theory, if the quality of information is good enough, near-perfect reliability can still be preserved in the face of such fluctuations

²⁷ If treaty members know ahead of time that capacity limitations are going to be a persistent problem, they should revise the agreement for the sake of efficiency. If they know that certain social or economic conditions will make it impossible for a state to comply with an agreement, they can describe these conditions in the agreement so that they do not count as "intentional" noncompliance—as they often do in the case of loans and in many broad and far-reaching agreements. There are limits to the extent to which this can be done, however, and once we are past those limits we can sensibly talk about shocks that can affect a state's reliability or reputation in equilibrium, at least if they are transient.

²⁸ There is a dangerous tendency in the literature to broaden the definition of "unintentional" to cover any defection that is remotely associated with a transient event. This can be a problem because each of the Chayeses's three factors involves an endogenous element that can easily lead to moral hazard. See Downs, *supra* note 4, at 331.

²⁹ Another source of indeterminacy with a similar effect springs from endemic rather than transient uncertainties associated with the cost of compliance. For example, developed states, much less developing states, do not know exactly how much it will cost to achieve different levels of compliance with complicated agreements like CITES or the World Trade Organization (WTO). In this cost-informational sense, the contract that each state makes is necessarily incomplete. Under these circumstances, it may also be more efficient to embrace a moderate level of punishment.

in utility by simply raising the penalty for defection. For example, states could collectively agree to respond to a defection by immediately abrogating the treaty forever and refusing to renegotiate it—a punishment known appropriately enough as the “grim strategy.” Assuming that a state’s discount rate is not extremely high, the cost that this represents should be sufficient to offset any transient benefit that a state can get by defecting when it experienced the kind of shocks described above.

Yet this is not the course that states pursue. On the contrary, the penalties specified in real-world treaties never resemble anything approaching the draconian proportions of the grim strategy and are characteristically relatively weak. In the history of international cooperation there are few if any instances where a state punishes a state that violates an agreement by abrogating the treaty and refusing to enter into a similar treaty. One might argue that this is a consequence of the fact that only modest penalties are needed to provide near-perfect compliance; therefore there is no reason to use the grim strategy. However, this cannot be the explanation because, as we have already seen, compliance, while often good, is usually less than perfect in any number of areas.

Another, more sophisticated, answer to the question of why states do not use stronger penalties to insure perfect compliance involves a slippery concept called “renegotiation proofness,” which puts a limit on the size of the punishment that can be used. To understand how it works, assume that a group of oil-producing states decides to maximize cooperation by collectively agreeing to punish overproduction on the part of any one state by jointly abrogating the treaty forever. Some time after the treaty is made, one state ends up increasing production and violating the treaty because of some unexpected series of events such as its leader having to raise money quickly to overcome a rebellion in one of its provinces. The other states respond by making good on their punishment threat and resume production at the level that existed before the agreement was signed. After a significant time period elapses, the president of the state that violated the treaty communicates to the other states’ leaders that, while she appreciates their mutual pledge to defect forever in the face of any violation, she can assure them that the period of unrestricted competition that has ensued has already cost her state more than it gained by breaking the treaty: “I and my state have already paid the price of our mistake. To continue producing at the competitive level will hurt you as well as us and it will accomplish nothing. Let us (renegotiate a) return to restricted production so that we will all be better off.”

This is the logic of renegotiation proofness. In prospect, whenever it is believed by the players that a punishment strategy would be vulnerable to such an offer, it is said not to be renegotiation proof and, as such, is judged not to be credible. As a practical matter, one can think of renegotiation proofness as restricting the size of the punishment to the largest value that

will never simultaneously be in both players' interest to reduce.³⁰ Unfortunately, the task of formally translating the above description into formal terms turns out to be more difficult than it might look, and none of the various versions in the literature is regularly employed.

Fortunately, the existence of stochastic compliance costs suggests an alternative explanation to renegotiation proofness for limiting the magnitude of the punishment that is both more intuitively attractive and easier to formalize. These costs do this by creating a situation where states are collectively better off with a treaty and an associated punishment regime that allows them to periodically defect than they are with a treaty that forces them to always cooperate. A penalty is still needed to prevent the actors from defecting exploitatively at every opportunity. However, the duration or severity of the penalty is adjusted so as to optimize the expected benefits of the treaty given the anticipated fluctuations of the cost and/or benefit structure. As a result, we have treaties in which states exhibit high but not perfect levels of reliability, and treaty reliability rates differ from one agreement to another because they are subject to different patterns of transient shocks. For example, scholars have argued that the weak enforcement norm of the General Agreement on Tariffs and Trade (GATT) was the result of the existence of the sort of transient and basically exogenous states of the world described above. Many (if not most) states did not want aggressive enforcement of GATT because they knew that there would be times when they themselves would find it advantageous to depart temporarily from a free-trade standard. This is because the political, if not the economic, benefit of free trade varies with changes in the state of the economy, particularly changes in unemployment in sectors that are import competitive. When unemployment is high, the domestic pressures to apply tariffs become strong, and politicians find that they are often better off—at least until the economy recovers—responding to those demands with temporary protectionist policies and paying the costs (both reputational and otherwise) of doing so.³¹

In theory, states could ignore the existence of the transient shocks and the

³⁰ For more extended discussions of renegotiation proofness, see Dilip J. Abreu, David Pearce, & Ennio Stacchetti, *Optimal Character of Equilibria with Imperfect Monitoring*, 39 *J. Econ. Theory*, 251 (1986); Jean-Pierre Benoit & Vijay Krishna, *Renegotiation in Finitely Repeated Games*, 61 *Econometrica* 303 (1993); B. Douglas Bernheim, Bezalel Peleg, & Michael D. Whinston, *Coalition-Proof Nash Equilibria I: Concepts*, 42 *J. Econ. Theory* 1 (1987); B. Douglas Bernheim & Debraj Ray, *Collective Dynamic Consistency in Repeated Games*, 1 *Games & Econ. Behav.* 295 (1989); Joseph Farrell & Eric Maskin, *Renegotiation in Repeated Games*, 1 *Games & Econ. Behav.* 327 (1989); David Pearce, *Renegotiation-Proof Equilibria: Collective Rationality and Intertemporal Cooperation* (typescript revision of 1987 Cowles Foundation Discussion Paper No. 855, Yale Univ., Dep't Econ. 1988); Eric Van Damme, *Renegotiation-Proof Equilibria in Repeated Prisoner's Dilemma*, 47 *J. Econ. Theory* 206 (1989).

³¹ Alan O. Sykes, *Mandatory Retaliation for Breach of Trade Agreements: Some Thoughts on the Strategic Design of Section 301*, 8 *B.U. Int'l L. J.* 301 (1990); George W. Downs & David M. Rocke, *Optimal Imperfection? Domestic Uncertainty and Institutions in International Relations* 88–89 (1995).

differences in reliability across treaties and act as if states possessed a single reputation. However, except in cases where a special complementarity exists in the utilities the states place on a few of the agreements, it would not be rational to enact a unitary reputation for the same reason that large reciprocal punishments fail. That is, it would be more efficient for both states if they did not presume a single behavior in disparate agreements.³²

One might object to this picture of multiple and relatively independent reputations by pointing to situations where states have responded to non-compliance in one area by retaliating in another, unrelated area (for example, sanctions against Iraq for violating the weapons inspection provisions of its agreement with the United Nations (UN)). However, we would argue that these generally represent a coercive linkage penalty rather than a true reputational penalty. A linkage penalty might appear to be equivalent to a reputational punishment because it is occurring as a consequence of a pattern of noncompliance, but it actually is something different. The states that are punishing the guilty state are not doing so because they anticipate that because the state has violated a human rights agreement it will be unreliable in connection with a trade treaty or fail to pay back its loans. Rather, states are simply trying to coerce the guilty state into changing its behavior in the same way that they might try to coerce a state into altering its behavior in an area where there was no treaty (for example, sanctions against South Africa during the waning years of apartheid).

Another way of seeing this is to consider that if the states inflicting the linkage penalty were worried about the defecting state's reliability, they would defensively reduce their level of cooperation in the area where they expected that the violator's next defection would hurt them (the victim state) the most. When a linkage penalty is involved, the states are offensively reducing their level of cooperation in the area where it will inflict the most damage on the violating state. The very rarity and selective nature of these linkages are testimony to their strategic nature. The United States has periodically threatened to oppose China's entry into the World Trade Organization (WTO) because of its human rights record, but it did not threaten simultaneously to oppose China's participation in other environmental or arms agreements, and the vast majority of states did not threaten to do anything at any time.³³

³² See Susanne Lohmann, *Linkage Politics*, 41 *J. Conflict Resol.* 38 (1997), for a discussion of the circumstances where linkages are possible. Under the best of circumstances, it is extremely unlikely that these conditions will ever be present in connection with more than a handful of agreements, and the existence of a stochastic cost structure would reduce the likelihood still further.

³³ Another sort of linkage behavior whereby a defecting state's reputation is downgraded in two areas such as trade and human rights following its failure to comply with its agreement in one area is theoretically possible (see *id.*). However, the aggregate effect of such linkage is also likely to be modest. This is partly because it requires a special complementarity in the utilities that states place on the linked agreements that will rarely be present and partly because

In an environment where compliance costs and benefits fluctuate stochastically, the existence of agreements with different values operate to multiply the number of reputations still further. That states do, in fact, attach different values to different agreements is obvious enough, but formal models of reputation rarely reflect this fact because it complicates the mathematics considerably and often distracts the reader from the point that they are trying to make. Still, for the purposes of understanding reputation it is worth considering the implications of the fact that such variation is the rule rather than the exception. The United States does not assign the same value to the North Atlantic Treaty Organization (NATO) that it does the South-East Asian Treaty Organization; nor does it assign the same value to WTO and a bilateral fishing treaty that it might have with a small state.

In multilateral agreements that regulate goods with relatively few externalities such as trade agreements, reputation will operate to protect large and powerful states the most—assuming that relationships with them are most likely to be highly valued by their treaty partner—and small states the least. In multilateral agreements that regulate public goods, defection from a given agreement affects every treaty partner the same, so the relative importance of states is largely immaterial. What matters in this case is how important the agreement is to the defecting state. If the agreement produces a high level of benefits, states will endure an appreciable rise in compliance costs before they defect—just as they will in connection with a valued relationship in the case of agreements that regulate private goods. If the agreement produces only modest benefits, it will take less of an increase in compliance costs before costs begin to outweigh benefits, making it desirable for the state to defect.

The efforts of newly established states or new regimes to establish good reputations will make the compliance rates of international agreements higher than they might be otherwise, as long as states have an expectation that the future agreements will, on average, be more valuable than their current ones. This is a requirement that is likely to be met in most cases because prospective partners are likely to “test” the reliability of such states in connection with agreements of relatively modest importance before agreeing to enter into agreements where defection is very costly.

Yet the impact of this strategizing on the overall level of compliance in the international system is likely to be quite modest. Such states make up only a small proportion of the total number of states at any given time, and their need to strategically “inflate” their reliability rate erodes relatively quickly. As these states enter more and more agreements and other states gather reliability information in connection with each successive agreement,

only a very large state or a group of states acting in a coordinated way will have the power to establish a linkage that changes another state’s behavior in a dramatic way. Under the best circumstances, the ability to establish an effective linkage will be negatively related to the size and power of the target state.

new regimes will be able to develop segmented or multiple reputations like other states at a modest cost.

IV. THE SIGNIFICANCE OF REPUTATION FOR INTERNATIONAL LAW

If our characterization of how reputation operates is correct, the reputational consequences of a given state's noncompliance with a particular agreement are real, but they are more narrowly bounded than the traditional theory predicted. Other states will revise their estimates of its reliability but only in connection with agreements that they believe (1) are affected by the same or similar sources of fluctuating costs and (2) are valued the same or less by the defecting state. The first tends to limit the reputational consequences of even sharp decreases in compliance with a given treaty to other treaties in the same area. This prediction is in keeping with the already existing tendency of a substantial minority of scholars to confine their discussion of reputational consequences to their impact on a specific regime.³⁴ It helps explain why, despite the prevalence of the unitary reputation assumption, examples of a state's defection from an agreement in one area (for example, environment) jeopardizing its reputation in every other area (for example, trade and security) are virtually nonexistent in the literature. More important, it helps account for why states often have widely divergent reputations in different areas.

This is not to deny that it can be a useful rhetorical convenience or an effective political strategy to speak of a state as if it had a single reputation—especially when the context is understood, for example, during a discussion of human rights—but it will be descriptively misleading if taken too seriously. Thus, the United States has one simple reputation for making good on its financial commitments with workers in the UN Office of the Secretary-General and another quite different simple reputation with officials of European states in connection with its financial commitments to NATO. Neither group is much concerned with characterizing the reliability of the United States in meeting its financial commitments in general. Those inside the Office of the Secretary-General are aware of the fact that the United States has paid its NATO bills, and NATO workers know that the United States is behind on its UN dues. However, they design their policies in response to the behavior of the United States in the subset of contexts that is relevant to them.

The claim that reputational inferences are localized or “segmented” helps explain why members of the North Atlantic community could have shown

³⁴ Whether the underlying logic of these authors who implicitly treat reputational implications as being confined to a specific regime is the same as that described here is unclear since they rarely give the reasons for their position. It is possible that some of them actually embrace the unitary perspective and only appear to believe in regime-specific reputations in the context of a given article whose focus is regime specific.

solidarity in dealing with Slobodan Milosevic's Yugoslavia at the same moment they were squabbling over the sale of bananas. While there was no shortage of dispositional-like statements in the United States about European Union "protectionism" and "lack of commitment to multilateralism," such judgments had few consequences for the ability of the allies to conduct the war. In this case, both the utility and stochastic elements of the trade and security treaties were not sufficiently related for reputational consequences of one treaty to affect the other treaty.

Treaties in the same specific area with similar stochastic costs may still have few reputational connections if the treaties are valued significantly differently. In particular, a state's behavior in connection with a treaty that it derives little benefit from does not change its reputation in connection with those that it clearly values much more. This serves to distinguish our position from that often found in the alliance literature. The latter suggests that states are prepared to reestimate each other's reputations at the drop of a hat and that, as a consequence, a state must be scrupulous about aggressively upholding every security commitment that it might have made. For example, when a small number of Haitian security forces turned away U.S. troops trying to land at Port-au-Prince as part of a UN mission, a U.S. policy maker is said to have remarked, "This embarrassment has dealt a blow to American credibility, one of our most precious assets."³⁵

While such statements are understandable in the heat of the moment, our model predicts that whatever the policy maker might have said in frustration, he did not rush to recommend to his superiors that they immediately call Germany to reassure officials there that their security umbrella was still intact. German officials did not need reassurance because it never would have occurred to them that United States valued the two commitments equally. It follows that from the perspective of our model, the argument that a failure of NATO in an area like Kosovo or Macedonia will seriously erode NATO's credibility in connection with its commitments to Western Europe is wrong for the same reason.³⁶

³⁵ Mercer, *supra* note 20, at 7.

³⁶ Mercer has generated the most elaborate empirical and theoretical critique of the traditional view of reputation; however, we are trying to explain somewhat different aspects of reputation. Mercer is interested in the reputation that states assign other states for resolve or "the extent to which a state will risk war to achieve its objectives" (*id.* at 2). We are interested in the reputational judgments states make with respect to the likelihood that other states will honor their treaty commitments in the future, a broader topic that has relevance to trade and environmental cooperation as well as to security commitments. Nevertheless, we agree with Mercer that states tend to have multiple and contradictory reputations (see *id.* at 32). The same is true of his contention that "actors base their predictions on the specifics of the situation or perhaps the history of a continuing relationship, and not on a state's aggregate behavior" (*id.* at 24). We are not surprised by the fact that "[l]ike their Soviet counterparts, American decision makers did not seem to infer that behavior in one area results in similar behavior in another area" (*id.*). Unlike Mercer, however, we do not see the existence of such behavior as a challenge to game theory so much as a challenge to the assumption of actor homogeneity that was made

The states that do have to worry about these sorts of commitment failures are smaller states that believe they are valued the same as the states that were abandoned. Of course, it is more complicated than this—the cost of fulfilling the commitment matters as well as the benefit of fulfilling it, and these costs might vary in such a way as to give the smaller state more or less confidence. Still, the principle that small states have the most to learn by paying attention to the noncompliance of their partners in connection to other relationships seems to be a general one. Because their relationship is likely to be viewed by the defector as being of the same or less importance, the chances that they too will be the victim of defection is fairly large. Large states do not have to worry as much about the noncompliance of a partner in connection with a third party spreading to them unless the third party is another large state. Otherwise, there is a reasonable likelihood that the defecting state values their relationship more.

The connection between the value of a relationship and the perceived opportunity costs associated with a contemplated defection has two effects on the extent to which reputation sustains international law. The first is perverse, if unremarkable. The fact that the reputational consequences of defecting from an important relationship are larger than those of defecting from a less important relationship means that reputation protects strong states more than weak states. This is really not so surprising. The reputational implications of a firm's violating a contract with its most important client are greater than they would be if the client were unimportant. However, the fact that reputation protects most those who require the least protection is still disconcerting.

The tendency of the magnitude of the reputational implications of a defection to be directly proportionate to the value of a relationship also implies that the contribution that reputation makes to sustain international law cooperation is greatest in connection with agreements that states think are the most beneficial.³⁷ Conversely, it has the least effect in connection with agreements that produce the smallest amount of benefits. This predicts that the average compliance rate will be somewhat higher in connection with relatively important agreements. It also may help account for why the quality of the compliance data that are available is so frequently related to the importance of the agreement.

From a reputational standpoint, the utility of the cooperation that an agreement represents or the opportunity cost of defecting from it is only half the story. Its reputational consequences are also a function of the extent to which the stochastic cost function that leads to defection from it is correlated with those connected with other regimes. While little is known about this, it seems

for convenience in models designed to look at issues differently than those that interest Mercer or us.

³⁷ This assumes that we are talking about agreements that are affected by the same stochastic cost function.

likely that trade agreements are quite “central” in this respect, since many of the shocks that affect trade agreements such as recessions also affect compliance with agreements in other areas. Security agreements also have a claim to centrality. Major shocks in that area are likely to affect trade agreements as well as human rights agreements.³⁸ Defections from environmental agreements, at least at the present time, seem to have more narrow implications for treaties in other areas, as have human rights treaties. Hence, their reputational consequences, at least in the rational choice sense, should be more restricted.³⁹ It follows, *ceteris paribus*, that reputation promotes compliance with international law most in trade and security and least in environmental regulation and human rights. The most important reputational consequences are those connected with the most important agreements in these areas.

The disconcerting implication that reputation fosters cooperation more in connection with large states than with small holds only for treaties that regulate the exchange of private or club goods where exclusion is possible such as trade agreements and alliances. If we consider public-goods agreements, the relative importance of states in determining the importance of a violation recedes into the background because it is no longer possible to defect against a particular state: any defection affects every state in the agreement. In this situation the contribution of reputation to promoting compliance and sustaining international law becomes a function of the importance of the agreement as a whole. The agreements that get the most reputational help are the agreements that states value the most and whose stochastic cost component is the most highly correlated with that of other agreements.

Other effects of reputation on compliance with international law are more marginal, but worth noting. Just as one would expect, new states and states with new regimes will on average be more sensitive to reputational consideration and hence more apt to comply with international law than the average state. This is because they have not yet had the opportunity to develop multiple reputations by signaling consistent variations in their reliability rates across agreements. However, if the incidence of compliance opportunities is relatively high, as it is in trade and human rights agreements, it will not take the newcomer long to establish multiple reputations. In an area like security where the opportunities to comply with a defensive alliance are sparser, multiple reputations will take longer to establish, and strategic behavior of all sorts will flourish longer. This fact might partially account for why reputation is such a popular topic for discussion in security circles and why there are so many disagreements about the reputational consequences of a particular act.

³⁸ Joanne Gowa, *Allies, Adversaries, and International Trade* (1994).

³⁹ The reputational damage with respect to a state’s social standing in the international community might of course be as great or greater.

V. CONCLUDING COMMENTS

The claims of this paper that the reputational consequences of most defections affect only a subset of current and future agreements may be interpreted as further support for the revisionist claim that reputation is irrelevant. This is not the case. To say that the power of reputation to enforce compliance is usually modest is no more a dismissal of reputation than the claim that many cooperative agreements are relatively shallow is a dismissal of cooperation. Reputation matters, just not so much as some might like.

Of course, the theory presented here is relatively simple. It is designed primarily to understand the boundaries of reputational inference and how they relate to two stylized facts of cooperative life not easily explained by most conventional theories of reputation: (1) states have different levels of reliability in connection with different agreements, and (2) there is considerable evidence that states possess multiple or segmented reputations. Inevitably, many other interesting issues have been ignored or mentioned only in passing.

One set of such issues is connected with the role of information and the capacity of states to process it. The model we have described here is very information intensive. To behave optimally, states not only have to keep track of the reliability rates connected with different agreements, but they have to estimate the degree of similarity among them with respect to the values states assign them and the factors that determine variations in their compliance costs. Even if we are correct in believing that states pay attention to these factors, their capacity for making distinctions may not be as refined as efficiency demands.⁴⁰

Information-processing issues are particularly important for understanding the impact of reputation on new states and states that have recently experienced wholesale changes in their form of government. Although the international system usually contains only a handful of these states, the impact of reputation on their fortunes may be far greater than for more established states. How quickly can these states develop multiple reputations, and what is the cost of not being able to do so immediately? Where are multiple reputations likely to first appear? Conversely, what are the reputational implications of the fact that compliance information for many agreements is extremely poor and often based on self-reporting?

Finally, the question of how reputation affects the fortunes of small and or developing states demands more attention because it contrasts so sharply with the picture of reputation leveling the playing field between the weak

⁴⁰ A more behavioral approach might also prove helpful regarding this issue.

and the strong. Not only does it appear that a defection against such states tends to have the fewest reputational consequences for the violator, but the reputational penalties that they pay for their own defections will tend to be quite large.