Article

JOURNAL OF THEORETICAL POLITICS

Depth versus rigidity in the design of international trade agreements

Journal of Theoretical Politics
2014, Vol. 26(3) 468–495

DOI:10.1177/0951629813505723
jtp.sagepub.com
SAGE

Leslie Johns
Department of Political Science, University of California, Los Angeles, USA

Abstract
I examine the impact of depth and rigidity in international trade agreements. Increasing the depth of required cooperation lowers the likelihood of full compliance and the stability of a trade regime. In contrast, increasing the rigidity of an agreement raises the likelihood of full compliance and lowers stability. Both depth and rigidity can lower tariffs if a state does not defect from its treaty obligations. I argue that if we control for the benefits of trade liberalization, then observable treaties will have a negative relationship between depth and rigidity. Deep agreements will be flexible, while shallow agreements will be rigid.

Keywords
International law; international organization; international trade

1. Introduction

States create international agreements to secure mutual gains from cooperation. Many examples of international cooperation—including international trade and military alliances—are club good problems. Members of the cooperative regime must implement policies that are individually costly in order to produce benefits that are shared by members of the cooperative regime. For example, the tariff reductions required by a trade agreement are politically costly for a leader since they limit his ability to protect import-competing industries. However, a system of tariff reductions increases overall social welfare by benefiting consumers and exporters. Similarly, military alliances require costly effort by individual states in order to provide the benefit of collective military protection for the alliance's members.

These strategic incentives create a free-rider problem since states do not fully internalize the benefits of their own actions. Individual states have incentive to shirk on their

Corresponding author:
Leslie Johns, Department of Political Science, University of California, 405 Hilgard Avenue, Los Angeles, CA 90095, USA.
Email: ljohns@polisci.ucla.edu

own effort as long as they can continue to benefit from the efforts of others. For example, political leaders are often tempted to violate trade concessions—by imposing safeguards, antidumping duties, and other trade restrictions—while still expecting to benefit from the trade concessions made by other states. The costs of cooperation are affected by domestic political and economic pressure that fluctuates over time. A country experiencing a severe recession or a leader in the midst of a tight reelection campaign has increased incentive to shirk on international trade obligations and protect domestic industries. Similarly, contributions to a military alliance benefit arms suppliers and the military at the cost of the general tax-payer. This inherent redistribution of wealth has political consequences for leaders seeking reelection.

International agreements must have two basic components. First, states must choose the depth of cooperation by specifying what constitutes cooperation. Second, states must decide how to treat members that violate their cooperative obligations during tough times; how rigid will the regime be in punishing leaders who violate their treaty commitments because of domestic pressure? I argue that the depth and rigidity of international treaties are rationally designed in order to maximize the benefits of cooperation. States must take into account the impact of treaty design on two outcomes when designing treaties. First, the design of a treaty affects full compliance—whether the actual behavior of the state conforms to behavior prescribed by the cooperative treaty. Second, depth and rigidity affect the stability of the cooperative regime—the ability of the regime to endure.

In order to explore these issues more closely, I focus on the design of international trade agreements. In these agreements states choose the depth of concessions by specifying tariff bindings. Lower tariff bindings require deeper cooperation. The design of dispute settlement provisions affects the rigidity of the agreement—the degree to which defections from prescribed behavior are tolerated. Less rigid agreements include flexibility provisions—such as safeguards and antidumping procedures—that allow states to sometimes violate their trading obligations without abrogating the treaty.

I show that the depth and rigidity of international trade agreements have complex effects on state behavior. Increasing the depth of required cooperation—by decreasing tariff bindings—lowers the likelihood of full compliance and the stability of a trade regime. In contrast, increasing the rigidity of an agreement—by reducing opportunities for tolerated defection—raises the likelihood of full compliance and lowers stability. Additionally, depth and rigidity can lower tariffs if a state does not defect from its treaty obligations. States must balance these effects—on full compliance, stability, and tariffs—when they design optimal treaties. I argue that if we control for the benefits of trade liberalization, then observable treaties will have a negative relationship between depth and rigidity. International trade agreements that mandate deep tariff reductions will be more likely to include flexibility mechanisms—such as safeguards and antidumping procedures—than agreements that require shallow cooperation.

2. Depth and rigidity in international trade agreements

Recent studies of international law and organizations emphasize the impact of treaty design on political and economic outcomes (Gilligan and Johns, 2012). One key element of treaty design is the depth of cooperation—the tariff bindings in a trade agreement.[1]

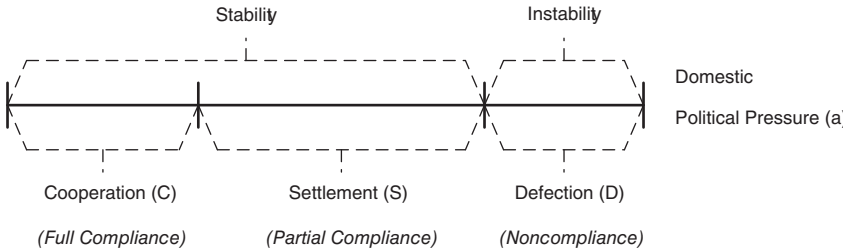

Figure 1. Compliance and stability in equilibrium.

Deep treaties create low tariff bindings, while shallower treaties create higher bindings that allow a state to choose higher tariffs without violating its treaty obligations.

A second key element is the rigidity of the agreement—the degree to which treaty violations are tolerated. Empirical scholars have demonstrated that trade agreements vary tremendously in their rigidity.[2] Consider safeguards, which allow a treaty member to violate a tariff binding if a surge in imports harms or threatens a domestic industry. Trade agreements vary in the criteria used to identify such import surges. Some treaties allow a single member to unilaterally identify surges, while others require consultations or specify objective standards that are monitored by a treaty body. Trade agreements also vary in the discretion they grant states to respond to import surges. Some treaties impose time-limits or explicitly require compensation to affected trade partners, while others impose no constraints whatsoever.[3] Similar variation exists in the design of other flexibility-enhancing devices, such as antidumping and countervailing duties.

The study of international agreements cannot be divorced from domestic politics (Goldstein and Martin, 2000). Sometimes a country will experience tough times, and a political leader will be pressured to protect domestic industries by violating trade obligations. As domestic political conditions shift, the leader's incentives to abide by the earlier concessions will change.

The primary objective of trade agreements is to promote trade by reducing tariffs. In the model below, I show that if a leader experiences low pressure to protect domestic industries, he will be more likely to meet his primary treaty obligation by choosing tariffs that are lower than the binding. I refer to this as 'cooperation' because the state has met all of its obligations under the cooperative regime. It is in full compliance with the agreement. This is shown in Figure 1.

However, trade agreements also specify secondary obligations by creating rules about how a state must behave when it violates its tariff bindings. For example, some treaties require only consultations, while others create elaborate dispute settlement procedures (DSPs) that allow for litigation before a third party. Sometimes a leader will be unwilling or unable to abide by a tariff binding. In the model below, if the leader faces moderate domestic political pressure, then he will violate the tariff binding, but abide by the treaty's DSP. I refer to this as 'settlement' because a state has violated its primary obligations, but abided by its secondary obligations. The state is in partial compliance with the trade agreement.[4] If the leader experiences high political pressure to protect domestic industries, then he will break the tariff binding and refuse to follow the treaty's DSP. Such 'defection' results in noncompliance with the agreement because the state has met none

of its treaty obligations. I assume that the agreement remains in effect if no state defects. I refer to this as 'stability'. However, if at least one state chooses defection, then there is 'instability' because states are not constrained by the agreement in future periods.

Previous research argues that depth lowers compliance rates (Downs et al., 1996). As an agreement requires deeper cooperation from states—more significant changes in behavior from what the state would do in the absence of the agreement—there will be less full compliance with its provisions. However, the relationship between depth and the stability of an institution has not been explored because this research has not examined settlement. Are agreements that require deeper cooperation more or less stable than agreements that demand less?

Many IR scholars argue that strong enforcement mechanisms are necessary to support deep international cooperation. Since states face short-term incentives to violate agreements, large punishments for noncooperative behavior—usually the exclusion of a violator from a treaty regime in future periods—are believed to be necessary to deter defection (Downs et al., 1996). However, other scholars argue that treaties will generate more cooperation in the long run if they allow violators to pay a fine in exchange for continued membership in the regime (Downs and Rocke, 1995; Rosendorff, 2005; Rosendorff and Milner, 2001).

I build on two models from this latter scholarship: Rosendorff and Milner (2001) and Rosendorff (2005). In these models, a leader experiences a stochastic shock to his cost of cooperation, and then chooses between a cooperative or a defection tariff. Absent a trade agreement, if the leader defects, he is punished by the collapse of cooperation in all future periods. In contrast, under a trade agreement, the leader can invoke an 'escape clause' that allows him to defect without punishment if he pays a fixed fine. Leaders who commit major violations can pay the same fine as leaders who commit only minor violations.[5] These earlier models show that an escape clause increases the stability of a trade agreement, and there always exists an escape clause that increases the utility of the players relative to their utility without an agreement.

In contrast to this earlier theoretical work, I treat the rigidity of an agreement as a continuous measure. Suppose that a temporary violation followed by a return to future cooperation comes with a cost. A rigid agreement imposes a high cost for a temporary violation, which leaves the leader with little policy discretion during tough times. In contrast, if the agreement is flexible, then there is little cost to remaining a member of the regime after temporarily violating a treaty. We can describe the rigidity of an agreement by how costly it is for a state to 'purchase forgiveness' from other treaty members when it temporary violates its primary obligations. Whereas earlier models compare state behavior with and without a trade agreement, my analysis allows for a more nuanced understanding of variation in the design of international law and institutions.

Additionally, I analyze the impact of the depth of tariff bindings, which these earlier models do not examine. I also analyze full compliance and tariffs, whereas earlier models focus exclusively on stability. Finally, I assume that fines vary in response to the size of violations: a leader who commits a minor violation is allowed to pay a lower fine than a leader who commits a major violation. This change adds nuance to dispute settlement because it allows a leader to adjust the size of his tariff violations (and hence the size of the associated fine) in response to domestic political pressure.[6]

My model shows that depth lowers the likelihood of full compliance and the stability of the treaty regime. If the trade agreement grows deeper, the leader is required to choose a lower tariff. Suppose that the leader is indifferent to both cooperation and settlement. If the leader cooperates, he must fully internalize the cost of the binding. However, if he settles, then he must only compensate a share of his partner's loss. As depth increases, settlement becomes more attractive than cooperation. Similarly, suppose that the leader is indifferent between settlement and defection. If the trade agreement grows deeper, then the size of the leader's violation grows larger, and the cost of settlement increases. Depth does not have the same impact on the defection payoff because the leader does not pay the fine. So as depth increases, defection becomes more attractive than settlement. From the perspective of Figure 1, depth narrows the cooperation interval and expands the defection interval. As an agreement grows deeper, its members are less likely to fully comply, and the agreement is less likely to survive in future periods.

I also show that rigidity decreases stability, but increases full compliance. If the trade agreement grows more rigid, the leader must pay a larger fine to remain within the agreement if he violates the tariff binding. Once again, suppose that the leader is indifferent between cooperation and settlement. An increase in rigidity does not affect the leader's payoff from cooperation, but it makes settlement more costly. So as rigidity increases, settlement becomes less attractive than cooperation. Finally, if the leader is indifferent between settlement and defection, an increase in rigidity makes settlement more costly, but does not affect the leader's payoff from defection. Settlement becomes less attractive than defection. Rigidity widens the cooperation and defection intervals in Figure 1. This suggests that the primary benefit of flexibility is that it increases the stability of the agreement.[7]

Both depth and rigidity can lower tariffs if the leader cooperates or settles. Suppose the leader is under very low political pressure and will cooperate, regardless of small changes in the design of the treaty. A decrease in the tariff binding will lower his tariff, and a change in rigidity will not impact his tariff because he is in full compliance. Alternatively, suppose the leader is under moderate pressure and will settle. An increase in either depth or rigidity raises the cost of settlement. To offset this change, the leader will lower his tariff. Depth and rigidity change not only the width of the intervals in Figure 1, but also the tariffs that are chosen by the leader within each interval.

Depth and rigidity have complex effects on state behavior. States must balance these effects—on full compliance, stability, and tariffs—when they design optimal treaties. As the final part of my analysis, I examine the attributes of treaty designs that maximize a state's ex ante expected utility. I show that if we control for the cooperative benefits that are achievable by treaty members, then observed treaties—those trade agreements that are written and signed by states—will have a negative relationship between depth and rigidity. Deep tariff bindings will be accompanied by flexibility provisions, while more rigid agreements will have shallower trade bindings.

3. Theory

I consider the strategic behavior of two states, which I call 'home' and 'foreign'.[8] The game has two stages. In the first stage, the states design the text of a cooperative treaty. To fully specify the treaty, they must specify the depth of cooperation and the rigidity of the

treaty in addressing violations. Depth of cooperation is represented by the tariff binding, t_B.[9] These bindings are the maximum tariffs that are permissible under the treaty. The rigidity of the treaty is represented by model parameter σ. Rather than fully modeling the many different forms of rigidity in trade agreements, I adopt a general model of the settlement of trade disputes. Given a tariff binding, we can calculate the losses that a state suffers when its trading partner violates the tariff binding. The rigidity parameter, $0 < \sigma < 1$, denotes the minimal share of these losses that must be compensated via a settlement in order for a state that has violated its binding to remain in partial compliance with the treaty. In the second stage of the game, states play an infinitely repeated cooperation game. In each period, leaders must simultaneously choose tariffs in response to stochastic domestic political pressure. If a state violates its tariff binding it must also choose how much (if any) compensation it will pay to its trading partner. This subgame is infinitely repeated.[10]

3.1. Settlement of trade disputes

Settlements compensate states that are harmed by their partner's trade policies. A treaty violation creates a gain for the violator, and a loss for its trading partner. A settlement is a mutually acceptable outcome that reduces this gain and mitigates this loss. A violator usually compensates its trading partner via policy changes—such as trade measure adjustments, technical assistance, or the withdrawal of equivalent concessions—rather than direct monetary payments or fines. Nonetheless, settlements redistribute the gains and losses that were created by the treaty violation in the first place. In this sense, we can conceptualize a settlement as a transfer from a violator to its trading partner. We cannot use this approach to examine the dynamics of bargaining and litigation between states after a violation has occurred. However, we can interpret settlement payoffs as the expected utility from a subgame in which states negotiate and litigate under the terms of a treaty text.[11] Alternatively, we can adopt an equilibrium-as-institution perspective and interpret σ as a shared understanding of appropriate behavior.[12]

The design of trade agreements affects the size of settlements in practice even though they rarely prescribe precise levels of compensation. Some agreements impose strict dispute settlement rules, while others are more permissive. These rules affect a member's perceptions about the likelihood and magnitude of future compensation for trading violations. The more rigid the rules, the higher the expected cost of breaking a tariff binding. For example, consider antidumping (AD) procedures. If a state believes that an imported good has been dumped (sold below its normal price), it can raise the price of this imported good by imposing an AD duty. This duty raises the effective tariff on the good above the tariff binding, which benefits the importer and harms the exporter. Trade agreements vary greatly in their AD restrictions. Some agreements allow states complete discretion in determining whether dumping has occurred, while others create complex technical guidelines. Some provide no oversight, while others establish monitoring institutions. In those agreements that create formal dispute settlement institutions, some restrict the institution to informal recommendations, while others empower the institution to issue legal rulings.[13] The more rigid the agreement, the more difficult it is for a state to impose an AD duty on an imported good. Even if a state's AD duty is permissible, it must still bear the cost of oversight and legal challenges.

While a settlement mitigates the gains and losses from a treaty violation, I assume that trading partners will never be 'made whole' ($\sigma < 1$). I make this assumption for both theoretical and empirical reasons. From a theoretical perspective, this assumption ensures that a state never profits when its trading partner violates the treaty. The structure of prisoner's dilemma payoffs is preserved, and my results are not driven by opportunities for efficient breach. Even if trading partners reach a large settlement, a state is still a 'sucker' if it cooperates while its partner does not.[14] This assumption is also desirable for empirical reasons. Many trade agreements create a legal standard of 'equivalent concessions' that ensures that injured states cannot secure more than the losses they have suffered. When this constraint is combined with the uncertainty inherent in dispute settlement and basic transaction costs—including the cost of litigation, bargaining, and delay—it is reasonable to think that a settlement will be a fraction of the damages incurred.

3.2. Cooperation subgame

Suppose that stage one of the game has ended and the states have specified the design of the trade agreement. Then the two states, home and foreign, play an infinitely repeated game in which they choose tariffs.[15] In each period, let t denote the tariff chosen by the home government, and τ denote the tariff chosen by the foreign government.[16] Each government cares about the welfare of its import-competing firms, exporters, and consumers. I use a reduced form model of the domestic economy so that I can focus on domestic politics and international cooperation. Consider the perspective of the home government. I assume that the benefit to an import-competing firm of the home tariff, t, is an increasing concave function, $u(t)$.[17] So import-competing firms always prefer higher tariffs on foreign goods, but their marginal benefit from a tariff declines as the tariff grows larger. Exporters care about the foreign tariff, τ. A higher foreign tariff creates lower foreign demand for the exporter's product, which leads to lower profits for the exporter. I capture this by assuming that the utility for home exporters from the foreign tariff is $-u(\tau)$. Finally, if the home government raises its own tariff, then home consumers experience higher prices. Consumers are unaffected by the foreign tariff. So I assume that an increase in the home tariffs will lead to a decline in political support from home consumers. I represent this loss by the linear term $-t$.

In every period, each state experiences political pressure from domestic economic actors. The level of this pressure is stochastic over time. I conceptualize this pressure as the weight that the government places on its import-competing industries relative to its own exporters and consumers in its political support calculation. Let a denote the political pressure experienced by the home government, and α denote the political pressure experienced by the foreign government.[18] So if the home government experiences high political pressure (draws a large value of a), then it has an increased incentive to raise its tariff. The political pressure experienced by each government is privately known to that government and is not observed by its trading partner. Both governments are uninformed about the pressure each might face in future periods. So the one-period utility functions of the home and foreign government—W and W^*, respectively—are as follows:

$$W(t, \tau, a) = a\,u(t) - t - u(\tau)$$
$$W^*(t, \tau, \alpha) = \alpha\,u(\tau) - \tau - u(t)$$

After observing its own current political pressure, each government simultaneously chooses a tariff.

If the foreign country violates the trade agreement by choosing a tariff that is too high ($\tau > \tau_B$), then the home government's losses are: $L(\tau) = W(t, \tau_B, a) - W(t, \tau, a) = u(\tau) - u(\tau_B)$. Under the terms of the treaty, the foreign state must pay $\sigma L(\tau)$ to the home country if it wants to settle. Similarly, if the home country violates the agreement by choosing $t > t_B$, then the foreign state's losses are: $L^*(t) = W^*(t_B, \tau, \alpha) - W^*(t, \tau, \alpha) = u(t) - u(t_B)$. The home country must pay $\sigma L^*(t)$ to the foreign country to settle the dispute. Decisions about whether to pay the required compensation are made voluntarily by each state. A treaty member can choose to violate the tariff binding and not pay compensation.

I assume that each state discounts its future payoffs according to the discount factor $0 < \delta < 1$. Additionally, I allow for the possibility that the benefits of cooperation may grow larger over time. Many scholars argue that a key factor driving international cooperation is the growth of pro-compliance domestic constituencies (Alter and Helfer, 2010; Gilligan, 1997; Simmons, 2009). While governments always face some pressure to protect import-competing industries, we might still believe that the benefits of trade cooperation can grow because of expansions in trade, reductions in the cost of transportation, or long-term changes in the allocation of resources across sectors.[19]

3.3. Equilibrium behavior in the cooperation subgame

As demonstrated in folk theorems, equilibria can generate many different behaviors in infinitely repeated games (Fudenberg and Tirole, 2000). For example, my model has an 'anarchic equilibrium' in which each state always chooses a 'defection tariff' that maximizes its one-period utility. In this equilibrium, a treaty does not change behavior: each state often violates the tariff binding and never compensates its trading partner.

Can states use institutions to escape this anarchy and promote international cooperation? I examine an equilibrium with three action profiles. First, a state can cooperate (C) by choosing a tariff that is lower than the tariff binding ($t \leq t_B$ and $\tau \leq \tau_B$). Second, a state can violate the tariff binding but then settle (S) by paying the required compensation. Finally, a state can defect (D) by violating the tariff binding and refusing to pay compensation.

A state will be tempted to violate the treaty when it faces high domestic political pressure. A state will only comply with the treaty if the punishment for defection is sufficiently large. Remember that a state can always maximize its one-period payoff by choosing the defection tariff. So the largest self-enforcing punishment for treaty defection is a return to anarchy for all future periods. This 'grim trigger' punishment creates more incentive for treaty compliance than all other punishment strategies, including 'tit-for-tat' or temporary reversions to anarchy. Using this punishment strategy allows us to examine the limits of institutional effectiveness.

In the model below, I assume that states impose a grim trigger punishment for defection: if at least one state violates the tariff binding and refuses to pay the required compensation, then all states abandon the treaty, revert to anarchy, and choose defection tariffs in all future periods. If neither player defects, then the treaty remains in effect in the next period. Nevertheless, all of my analytical results hold under alternative punishment strategies that create less incentive for treaty compliance.[20]

In the Appendix, I fully characterize the Bayesian Nash equilibrium of this game. Each leader will choose whether to comply (C), settle (S), or defect (D). This decision will be affected by the political pressure in the period. I identify two endogenous thresholds, a_S and a_D, that create three regions in the support of the political pressure variable, a, which I refer to as the 'type' of government.

Proposition 3.1. *There exists a Bayesian Nash equilibrium in which low types ($a < a_S$) comply (C), moderate types ($a_S \leq a \leq a_D$) settle (S), and high types ($a_D < a$) defect (D).*[21]

The first implication of my model relates to equilibrium tariffs.

Lemma 3.2. *If the leader chooses settlement or defection, then his tariff is increasing in his domestic political pressure.*

Figure 2 shows the equilibrium graphically. The horizontal axis is the domestic political pressure on the leader. The vertical axis is the tariff. The horizontal line at t_B indicates the tariff binding that is mandated by the trade agreement. Suppose that the leader faces low political pressure. If the leader were to choose his optimal tariff without regard to the state's international obligations, then the tariff that he would choose, t_D, is lower than the treaty binding.[22] When the leader faces little to no pressure from import-competing industries, his trade policy is unconstrained by the treaty and he complies with the binding. However, as political pressure increases, the leader faces an enhanced incentive to increase his tariff on the imported good. For sufficiently high political pressure, the leader's unconstrained ideal tariff (t_D) breaches the binding. In order to stay in compliance with the treaty, the leader will need to restrict its tariff to t_B. This is indicated by the flat portion of the bold line. As political pressure grows even larger, the political benefit of applying a tariff above t_B and then compensating his trading partner outweighs the benefit from cooperation. The optimal tariff in these circumstances, denoted by t_S, rises with the political pressure, but always lies below the defection tariff (t_D). Finally, if political pressure becomes very large, then defection becomes the optimal action. The leader will choose the defection tariff, but will not compensate his partner.

3.4. Compliance and stability in the cooperation subgame

To see how depth and rigidity affect the likelihood of full compliance and stability, we must investigate the effects of depth and rigidity on the equilibrium thresholds a_S and a_D. A rise in threshold a_S means that the likelihood of full compliance increases; a leader is more likely to choose tariffs at or below his binding. A rise in threshold a_D means that a leader is less likely to defect from the agreement, and the trading regime is more stable.

My first comparative static result concerns the likelihood of full compliance (C), which occurs when domestic political pressure is low ($a < a_S$).

Proposition 3.3. *Increasing the rigidity of the agreement increases the probability of full compliance. However, increasing the depth of the treaty—by choosing a lower tariff binding—lowers the probability of full compliance.*

Increasing the rigidity of the treaty means that if state violates its tariff obligation, then it must pay a larger settlement in order to ensure a return to cooperation in future

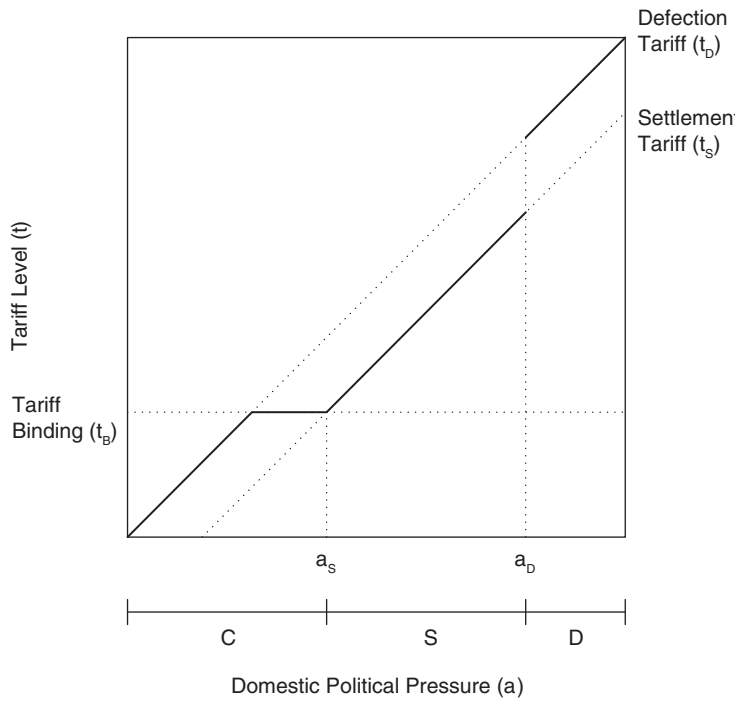

Figure 2. Equilibrium behavior.

periods. This makes each state less likely to violate in the first place. A lower tariff binding requires deeper cooperation from treaty members. Full compliance is made more difficult because more is required from states. This means that a state is more willing to commit a violation, be it tolerated or otherwise.

The impact of rigidity on full compliance may lead to the conclusion that stronger treaties—with little flexibility or opportunity for tolerated violations—are always a good idea. My next result shows that this conclusion is erroneous. Both depth and rigidity affect the threshold a_D, which determines the stability of the trading regime.

Proposition 3.4. *Increasing the growth in cooperative benefit over time increases the stability of the trading regime. However, increasing the depth or rigidity of the agreement decreases stability.*

If the growth in the benefit from trade cooperation increases, then leaders who defect from their tariff obligations become more willing to pay settlements rather than defecting and forgoing the future benefits of cooperation. This increases the stability of the trade agreement. However, if the tariff binding is lowered, then the treaty grows deeper and the agreement demands more from member states. Unsurprisingly, this means that a leader is more willing to forgo the future benefits of trade cooperation by defecting. The set of political pressure values for which the leader defects increases. Similarly, increasing the rigidity of the treaty has negative consequences for the stability of the regime. By raising the implicit price of tolerated defection, leaders that face high pressure to violate

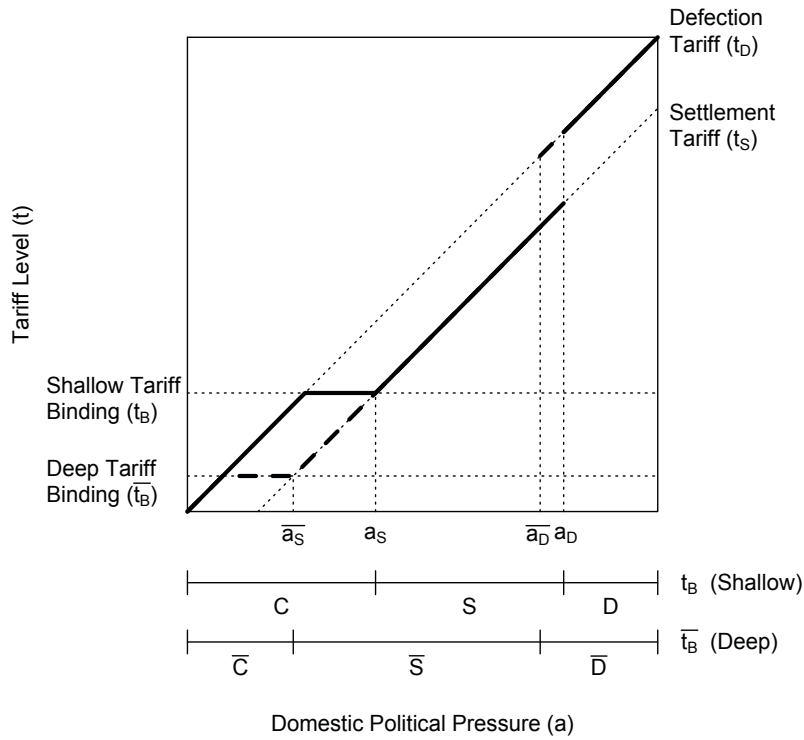

Figure 3. The impact of deeper tariff bindings.

the tariff binding will find it more beneficial to defect than to buy forgiveness by settling. Both depth and rigidity reduce the stability of the agreement.

The results from Propositions 3.3 and 3.4 are demonstrated graphically in Figures 3 and 4, which also show the impact of depth and rigidity on equilibrium tariffs. In Figure 3, I show the effect of lowering the tariff binding from t_B to $\overline{t_B}$. This makes the agreement more demanding on the member states because deeper cooperation is required by the treaty. The solid line denotes equilibrium tariffs under the original shallow binding t_B. The dashed line shows equilibrium tariffs under the deep binding, $\overline{t_B}$. As stated in Proposition 3.3, the lower threshold decreases from a_S to $\overline{a_S}$, which ensures that the region of full compliance shrinks from C to $\overline{C}$. Similarly, as the treaty grows deeper, the upper threshold shifts to the left from a_D to $\overline{a_D}$. This is the effect highlighted in Proposition 3.4. Increasing the depth of the treaty makes defection more likely. This reduces the stability of the agreement. Trade agreements that demand deeper cooperation get less compliance and have a higher risk of collapse. However, note that depth decreases the equilibrium tariff if the leader does not defect.

In Figure 4, I show the impact of increasing the rigidity of the trade agreement from σ to $\overline{\sigma}$. The solid line denotes equilibrium tariffs under a flexible agreement, σ. The dashed line shows equilibrium tariffs under the more rigid agreement, $\overline{\sigma}$. Rigidity reduces the settlement tariffs from t_S to $\overline{t_S}$. This follows naturally: if a state is going to be more severely penalized for a violation, then it will violate less by choosing a lower tariff.

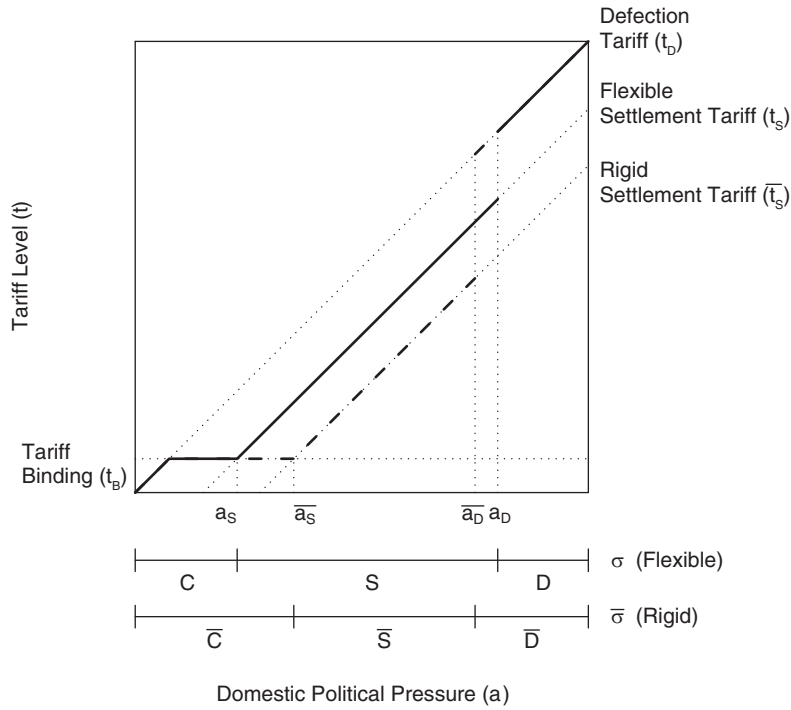

Figure 4. The impact of more rigidity.

Increased rigidity affects the likelihood of full compliance and stability by changing the two threshold values. The zone of full compliance increases in size from C to $\overline{C}$ because the lower threshold a_S rises to $\overline{a_S}$. However, the upper threshold shifts down from a_D to $\overline{a_D}$ as rigidity increases. This increases the region in which states defect. So rigidity reduces the stability of the regime. Additionally, the settlement region shrinks, which reduces the likelihood of partial compliance. So rigidity increases the likelihood of full cooperation, but decreases stability.

3.5. Depth versus rigidity in bargaining outcomes

We can be confident that these theoretical results are robust because the model structure is very general, but this generality comes with a cost: I cannot explicitly characterize the optimal depth and rigidity. However, I can analyze attributes of optimal treaties. I show that, when leaders design optimal treaties, they must trade off depth and rigidity.

When leaders design a treaty, they know that in the future they will need to choose trade policies that satisfy the competing interests of consumers, exporters, and import-competing industries. I assume that leaders know that future political pressure from these groups will fluctuate in response to exogenous shocks, but do not know the precise pressure that they will face in future periods. Leaders must design a treaty from behind a Rawlsian 'veil of ignorance' (Johns, 2012). While each leader differs in his preferred tariff after he signs the treaty and knows how much political pressure he faces, leaders have

identical ex ante preferences over the design of the trade agreement if they are symmetric; if they have the same beliefs about the distribution of shocks, patience, opportunities for growth, and utility functions.

As with any modeling choice, the assumption that the two leaders are symmetric—and hence have identical ex ante preferences over treaty design—has both drawbacks and benefits. The major drawback is that we cannot examine distributional conflicts. Suppose that leaders differ in their beliefs about future domestic political pressure. For example, the leader of a developed country may expect less pressure to protect import-competing industries than the leader of a developing country. This difference could cause leaders to differ in their preferred design of the trade agreement. We would only be able to examine the attributes of observed treaties if we made assumptions about the bargaining process. This is problematic because international negotiations lack fixed and well-specified bargaining protocols (Johns, 2007; Milner and Rosendorff, 1996). The major benefit of assuming symmetry is that we can examine the attributes of optimal treaties without being constrained by assumptions about the bargaining process. If leaders have identical ex ante preferences over the design of the trade agreement, then we can reduce any bargaining game into a decision-theoretic framework in which a leader maximizes her utility from the trade agreement by choosing optimal levels of depth and rigidity. This allows us to be confident that the model results are not being driven by arbitrary assumptions about the bargaining process; they are robust across all possible bargaining games.

There will always be a limit to the benefits that are 'achievable' via a trade agreement. From both a theoretical and empirical perspective, we must control for this variation in achievable benefits if we want to understand treaty design. For example, as a leader benefits more from trade liberalization, he is more likely to prefer an agreement that is deep and rigid. We might then conclude that the deeper the agreement, the more rigid it will be. However, we would be ignoring the fact that this relationship between depth and rigidity is being driven by a selection effect: leaders who like lower tariffs will select into treaties that are deeper and more rigid. I only examine treaty designs that maximize state utility when the level of achievable benefits is held fixed. This means that I have theoretically controlled for factors that drive a leader's preference over tariffs. This approach ameliorates selection bias. By only analyzing those treaty designs that maximize state utility, we are only making claims about those designs that are observable by researchers.[23]

Leaders who benefit greatly from expanded trade are likely to design deep and rigid treaties, while leaders who gain little from trade will prefer design shallow and flexible treaties. However, when we control for such variation in achievable benefits, there will be a negative relationship between depth and rigidity.[24] If a leader increases one of these treaty attributes, then he must decrease the other to maintain his optimal tariffs.

Proposition 3.5. *If we control for the level of achievable benefits, there will be a negative relationship between depth and rigidity in observed trade agreements—increased depth must be accompanied by decreased rigidity, and vice versa.*

Figure 5 illustrates the trade-off between depth and rigidity. Consider panel (a). Recall that if a leader experiences high domestic political pressure, he chooses the defection tariff. As the tariff binding grows deeper, the leader becomes more likely to defect, which increases tariffs. If a leader experiences low political pressure, he will remain a member of the treaty regime. As the tariff binding grows deeper, he will choose lower tariffs. The

latter effect outweighs the former, and deeper tariff bindings decrease the average tariff. In contrast, panel (b) shows that less rigidity increases the average tariff. As the rigidity of an agreement decreases, the leader is less likely to defect when he experiences high political pressure. He will choose settlement tariffs, rather than defection tariffs, which decreases tariffs. However, if the leader experiences low political pressure, he will be more likely to violate the tariff binding, which increases tariffs. Once again, the latter effect outweighs the former, and less rigidity increases the average tariff.

The optimal levels of depth and rigidity can vary across agreements.[25] Leaders will vary in how much they can benefit from a trade agreement. Those who benefit from low tariffs will design deep and rigid agreements, while those who prefer higher tariffs will design shallow and flexible agreements.[26] However, if we control for the level of achievable benefits, then there will be a negative relationship between depth and rigidity in the set of observable treaties.

By assuming that leaders design the treaty behind a Rawlsian 'veil of ignorance', I do not explicitly model the impact of domestic interest groups on the design of trade agreements.[27] For example, Kucik (2012) argues that exporters and consumers—who want open markets abroad and low domestic prices on goods, respectively—prefer deep and rigid agreements, while import-competing industries—who want closed markets at home—prefer shallow and flexible agreements. Each group's ability to lobby for its preferred outcome is affected by political factors, such as industry concentration, resources, and political mobilization. Nonetheless, leaders must satisfy these competing domestic interest groups, which suggests that leaders must balance depth and rigidity to maintain broad political support.

4. Empirical support for the theoretical model

Empirical tests of the theoretical relationships above are beyond the scope of this paper. Nonetheless, previous empirical studies support some parts of my theory. For example, Kucik and Reinhardt (2008) examine the impact of rigidity on the stability of trade agreements. They ask whether domestic AD procedures affect membership in a trade regime. They show that countries with domestic AD procedures are more likely to join and remain a member of the WTO than countries without these procedures. Additionally, they find that new members of the WTO are more likely to adopt AD procedures than non-members. This provides some modest evidence that rigidity decreases stability (Proposition 3.4). This study does not examine the impact of the depth of tariff bindings on the stability of the trading regime. To my knowledge, no empirical studies explore the relationship between depth and stability.

We also lack empirical analyses of the impact of either depth or rigidity on full compliance with trade agreements (Proposition 3.3) or chosen tariffs. Rational choice IR scholars generally presume that depth lowers compliance rates (Downs et al., 1996). Two recent studies examine the impact of International Monetary Fund (IMF) regulations on current account restrictions, and show that a deeper treaty commitment (voluntary acceptance of Article VIII of the IMF Agreement) is positively associated with IMF compliance (Simmons, 2000; Von Stein, 2005). These authors disagree about whether the treaty constrains states, but both argue that those countries that select into deep commitments are different from countries that do not.[28]

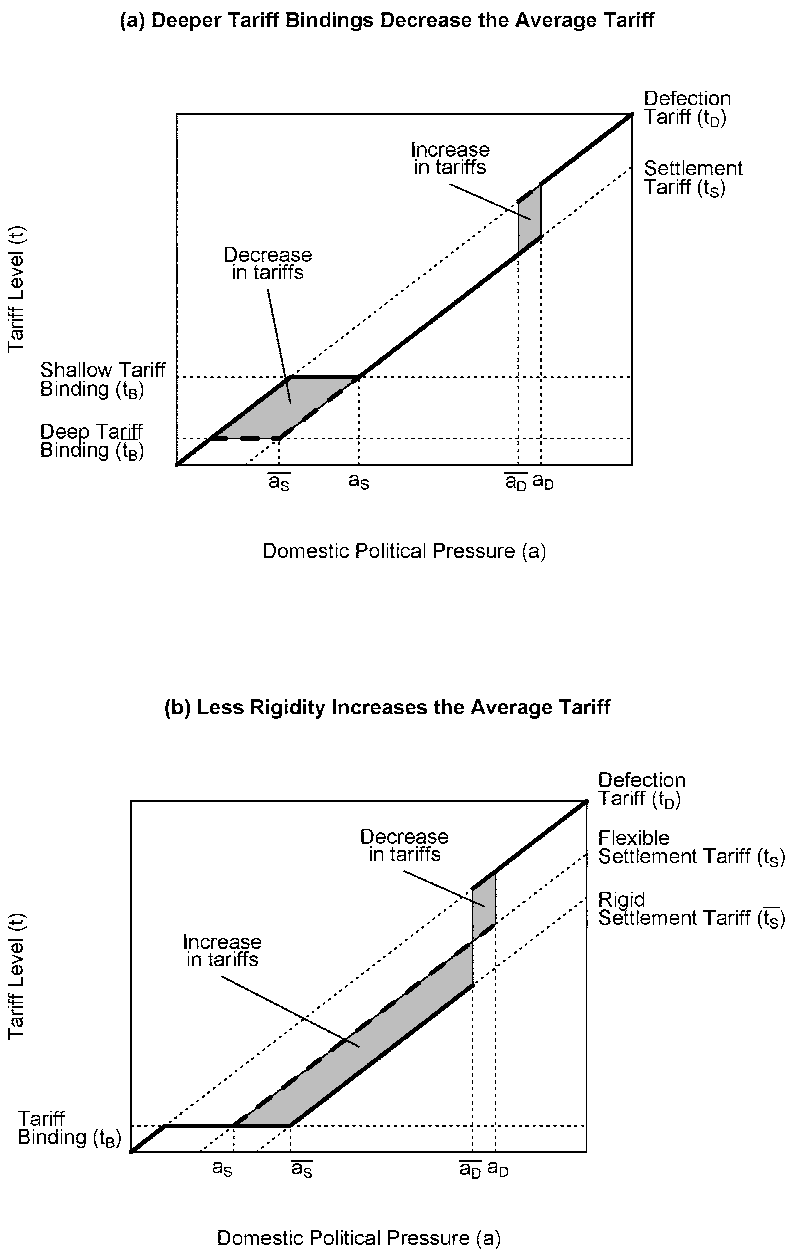

Figure 5. Depth versus rigidity.

This highlights the general difficulty of empirical analysis of international law: treaty membership and terms are endogenous to the treaty's expected effect on international cooperation. Direct empirical tests of the theoretical impact of depth and rigidity on full

compliance and stability are extremely difficult because states do not randomly select into treaties. My model suggests that any empirical analysis of international law should control for both the economic and political factors that shape the design of trade agreements ('selection' into treaty terms), as well as the subsequent impact of the treaty on economic outcomes. Alternatively, we must seek out natural experiments in which there is an exogenous shock to the terms of international cooperation. I provide an alternative approach to these standard methods. I assume that treaties are endogenous—that states design optimal treaties—and then identify an observable empirical relationship that is consistent with this endogeneity. This yields Proposition 3.5: if we control for the level of achievable benefits, there will be a negative relationship between depth and rigidity in observed trade agreements.

Recent empirical scholarship has examined political and economic factors that shape the design of dispute settlement procedures. While these studies focus on explaining treaty design, many reveal interesting empirical relationships among treaty attributes. Smith (2000) develops a detailed taxonomy and dataset of DSPs in regional trade agreements. He argues that economic interdependence, relative economic power, and the depth of integration all affect state preferences over treaty design. Also, he argues that trade agreements with more legalized DSPs are more effective because they place greater constraints on leaders. His study supports Proposition 3.5 by showing that trade agreements that require deeper integration are more likely to include flexibility-enhancing devices, such as third-party review of state disputes.

Similarly, Haftel (2012, 2013) constructs a dataset of 28 regional economic organizations and examines the influence of economic interdependence on institutional design. He codes many measures of the depth of economic cooperation, including: monetary and fiscal cooperation, sectoral harmonization, economic development, and provisions on the free movement of goods, services, capital, investment, and labor. His analysis shows that deeper cooperation is associated with decreased rigidity.

Finally, Baccini et al. (2012) argue that the relative power of domestic interest groups (as proxied by intra-industry trade) shapes the design of trade agreements. They construct a new dataset and use sophisticated methodological techniques to control for the fact that interest groups can simultaneously affect the depth and rigidity of a trade agreement. They present many complex findings about the impact of lobbying on treaty design. Most importantly from our perspective, they show that there is a negative correlation between depth and rigidity in observable treaties.

The endogeneity of treaty design to expected political and economic outcomes also makes it difficult to illustrate my theoretical arguments using case studies or qualitative evidence. Nonetheless, the trade-off between depth and rigidity in the design of trade agreements can be seen if we examine a natural experiment: India's trade policies during the 1990s.[29]

In the summer of 1991, India requested financial assistance from the IMF. The IMF often imposed 'conditionality' on loan recipients during this time period: the IMF would only provide funds if the recipient agreed to implement specific free market reforms. For example, the IMF often required loan recipients to open capital markets, reduce government spending, and privatize government-owned industries.[30] However, the IMF rarely mandated trade policy reforms because these were believed to fall under the purview of the GATT, not the IMF.

India's loan in 1991 came with a surprising condition: the IMF required India to liberalize its trade policies. The impact was sudden and dramatic. As documented by Hasan et al. (2007: 467): '[India's] mean tariff went from 128% before July 1991 to 94% in February 1992, 71% by February 1993, 55% in February 1994, and to roughly 35% by 1997–1998'. This collapse in tariff rates was so unexpected that many empirical scholars argue that India experienced an exogenous economic shock to its trade policies, and treat India's economic performance in the 1990s as a natural experiment (Goldberg et al., 2010; Hasan et al., 2007; Krishna and Mitra, 1998).

As tariff rates fell, India protected its domestic industries by dramatically expanding its use of safeguards and AD duties. Prior to the 1991 loan, India had never imposed a safeguard or an AD duty (Bown, 2012). India had no domestic safeguard procedures at the time. It did create domestic AD procedures in 1985, but had never actually used them to impose an AD duty (Bown and Tovar, 2011). Once tariff rates began to collapse, India created domestic safeguard procedures and began to aggressively impose safeguards and AD duties. Bown and Tovar (2011: 115) show that this relationship is statistically significant at the product level: the steeper the decline on a product's tariff, the more likely the product was to receive protection via a safeguard or an AD duty. They conclude that 'India unwound its commitment to reduce tariffs through use of AD and safeguard protection'. Simply put, India reduced the rigidity of its trade practices because it was unexpectedly compelled by the IMF to deepen its effective tariff concessions.

India rebalanced its depth and rigidity via unilateral changes in trade policy, rather than bilateral changes in treaty design. I do not explicitly model such a unilateral shift, yet the logic of my argument extends naturally to this case. India experienced an exogenous shock to the depth of its tariffs, but its trading partners did not. So the pressure to reduce the rigidity of trade obligations was unilateral, rather than bilateral.

5. Conclusion

International trade agreements are designed to promote international cooperation by lowering trade barriers and increasing the flows of trade among member states. However, in order for an agreement to be effective, it must be neither too strong nor too weak. I show that increasing the depth of cooperation lowers the likelihood of full compliance and the stability of the regime. In contrast, as the rigidity of an agreement increases, the likelihood of full compliance with the treaty increases while the stability of the regime decreases. I argue that if we control for the cooperative benefits that are achievable by treaty members, then observed treaties—those trade agreements that are written and signed by states—will have a negative relationship between depth and rigidity.

These results have many implications for the study of international cooperation, institutions, and law. First, the mechanisms driving the trade-off between the depth of cooperation and the rigidity of dispute resolution hold for a large set of cooperative problems. My results are driven by the conceptualization of international trade cooperation as a problem of club goods provision. While members of the treaty regime can be excluded from the benefits of cooperation—such as exclusion from a trade regime or security alliance—each state is tempted to cheat on its individual obligations since it does not fully internalize the benefits of its cooperative effort. Each state prefers to free-ride

on the efforts of others, particularly when it is under intense domestic pressure to violate its obligations. Rigid enforcement systems can ameliorate this pressure and enhance full compliance, but such systems also decrease stability by enhancing the likelihood of defection during tough times.

Second, my study challenges the reduction of legalization into a one-dimensional attribute, varying from 'hard' to 'soft' law (Abbott and Snidal, 2000; Goldstein et al., 2000). Hard law is defined as a system of precise rules with strong delegation to third parties and high obligation, while soft law scores low on each of these attributes. My analysis suggests that there is much interesting variation within hybrid, or mixed, legal systems that lie between these two extremes. As the arguments above illustrate, states must balance design attributes across multiple dimensions of a legal system. For many areas of law the richest theoretical and empirical accounts are likely to arise from exploring variation among treaty design attributes.

Third, a basic tenet of political science is that political institutions affect policy outcomes. However, recent empirical studies of international institutions find mixed evidence on this point. For example, Andrew Rose caused much consternation with his recent claim that 'membership in the GATT/WTO is *not* associated with enhanced trade' (2004a: 98). He has also provided evidence that the GATT/WTO has no impact on trade policies or volatility (Rose, 2004b, 2005). Similarly, Von Stein argues that Article VIII of the IMF, which forbids states from restricting their current accounts, 'appears to have little constraining power independent of the factors that lead states to sign' (Von Stein, 2005: 612). The details of these studies have been challenged by other empirical scholars.[31] Nevertheless, they provoke distress for many: why do international institutions often appear to be ineffective in changing state behavior?

Past theoretical accounts emphasize two factors that limit the effectiveness of international institutions: selection effects and exit options. Since a state chooses whether to be a member of an international institution, 'most treaties require states to make only modest departures from what they would have done in the absence of an agreement' (Downs et al., 1996: 380). Treaty regimes may provide screening effects that allow some states to cooperate while others are excluded (Von Stein, 2005). However, states are unlikely to select into institutions that significantly constrain their behavior. Even if a state does select into an institution, threats to leave the institution—by choosing an exit option—affect behavior within the institution. For example, the ability of states to leave the institution can limit cooperation and affect the distribution of costs and benefits within the membership (Johns, 2007; Voeten, 2001).

My theory offers an alternative and complementary explanation: treaties that ostensibly require deep cooperation are more likely to include flexibility mechanisms that permit temporary noncompliance. In essence, international institutions only require changes in behavior when leaders are experiencing good times at home. Treaties are designed to permit reversions to noncooperative behavior when leaders face tough times at home. This means that the impact of institutions on international cooperation is a function of domestic politics. Changes in domestic political and economic conditions can influence the efficacy of international institutions.

Finally, my theory highlights that advocates of international cooperation are not just limited to promoting deeper treaties. Another way that trade liberalization and other forms of cooperation can be enhanced is through the tightening of escape valves. In

essence, this is what occurs in the Dispute Settlement Body of the WTO. Advocates of liberalization promote restrictive and rigid interpretations of the escape mechanisms in the WTO treaty, while opponents push for more expansive and flexible interpretations of the treaty. This suggests that an understanding of international adjudication is key to a broader understanding of international cooperation.

Appendix

Let χ_N denote the continuation payoff if no treaty is in place. So χ_N is the reversion payoff from both players adopting the defection tariff in every time period. Let χ_C denote the continuation payoff if the treaty remains in effect. Suppose that there is growth in the value of the cooperative regime over time, where $\beta \geq 1$. If $\beta > 1$, then the values of χ_C and a_D will change over time. I suppress these time subscripts below without loss of generality. Recall that $\sigma \in [0, 1]$, $a, \alpha \sim_{iid} U[1, A]$ for large A, $u' > 0$, and $u'' < 0$.

Proof of Lemma 1. The home country's expected utility from violating the binding and not paying compensation (defection) is:

$$EU(D|t,a) = a\,u(t) - t - \int_1^A u(\tau(\alpha))\,dH(\alpha) + \int_{\alpha_S}^{\alpha_D} \sigma L\left(\tau\left(\alpha\right)\right) dH(\alpha) + \delta\,\chi_N$$

So the optimal defection tariff solves:

$$\begin{aligned} \frac{\partial EU(D|t,a)}{\partial t} &= a\,u'(t) - 1 = 0 \\ &\Leftrightarrow u'(t) = \frac{1}{a} \quad \Leftrightarrow \quad t_D(a) = u'^{-1}\left(\frac{1}{a}\right) \end{aligned}$$

This violates the binding iff:

$$t_D(a) = u'^{-1}\left(\frac{1}{a}\right) > t_B \quad \Leftrightarrow \quad \frac{1}{a} < u'\left(t_B\right) \quad \Leftrightarrow \quad a > \frac{1}{u'\left(t_B\right)} \equiv a_B$$

The home country's expected utility from violating the binding and paying compensation (settlement) is:

$$\begin{aligned} EU(S|t,a) = a\,u(t) - t - \sigma L^*(t) - \int_1^A u(\tau(\alpha))dH(\alpha) + \int_{\alpha_S}^{\alpha_D} \sigma L(\tau(\alpha))dH(\alpha) \\ + H\left(\alpha_D\right)\delta\beta\,\chi_C + \left[1 - H\left(\alpha_D\right)\right]\delta\,\chi_N \end{aligned}$$

So the optimal settlement tariff solves:

$$\begin{aligned} \frac{\partial EU(S|t,a)}{\partial t} &= a\,u'(t) - 1 - \sigma u'(t) = 0 \\ &\Leftrightarrow u'(t) = \frac{1}{a-\sigma} \quad \Leftrightarrow \quad t_S(a) = u'^{-1}\left(\frac{1}{a-\sigma}\right) \end{aligned}$$

This violates the binding iff:

$$t_S(a) = u'^{-1}\left(\frac{1}{a-\sigma}\right) > t_B \;\Leftrightarrow\; \frac{1}{a-\sigma} < u'(t_B)$$
$$\Leftrightarrow a > \frac{1}{u'(t_B)} + \sigma \equiv a_S$$

Note that: $t_S(a) < t_D(a)$ for all a. The optimal cooperative tariff is:

$$t_B(a) = \begin{cases} t_D(a) & \text{if } a < a_B \\ t_B & \text{if } a_B \leq a \end{cases}$$

□

Proof of Proposition 3.1. The home country's expected utility from actions C, S, and D given tariff levels from Lemma 3.2 are:

$$EU(C|t_B(a), a) = a\,u(t_B(a)) - t_B(a) - \int_1^A u(\tau(\alpha))dH(\alpha) + \int_{\alpha_S}^{\alpha_D} \sigma L(\tau(\alpha))dH(\alpha)$$
$$+ H(\alpha_D)\,\delta\beta\chi_C + [1 - H(\alpha_D)]\,\delta\chi_N$$
$$EU(S|t_S(a), a) = a\,u(t_S(a)) - t_S(a) - \sigma L^*(t_S(a)) - \int_1^A u(\tau(\alpha))dH(\alpha)$$
$$+ \int_{\alpha_S}^{\alpha_D} \sigma L(\tau(\alpha))dH(\alpha) + H(\alpha_D)\,\delta\beta\chi_C + [1 - H(\alpha_D)]\,\delta\chi_N$$
$$EU(D|t_D(a), a) = a\,u(t_D(a)) - t_D(a) - \int_1^A u(\tau(\alpha))dH(\alpha) + \int_{\alpha_S}^{\alpha_D} \sigma L(\tau(\alpha))dH(\alpha) + \delta\chi_N$$

To compare utility from actions C and S, define for $a_S \leq a$:

$$\hat{\Delta}(a) = EU(C|t_B(a), a) - EU(S|t_S(a), a)$$
$$= a\,u(t_B) - t_B - a\,u(t_S(a)) + t_S(a) + \sigma L^*(t_S(a))$$

Note that $t_S(a_S) = t_B$, so $\hat{\Delta}(a_S) = 0$. Also:

$$\frac{\partial \hat{\Delta}}{\partial a} = u(t_B) - u(t_S(a)) - (a - \sigma)\,u'(t_S(a))\,\frac{\partial t_S(a)}{\partial a} + \frac{\partial t_S(a)}{\partial a}$$
$$= u(t_B) - u(t_S(a)) < 0$$

So S strictly dominates C for all $a_S < a$. To compare utility from actions S and D, define for $a_S \leq a$:

$$\begin{aligned}
\bar{\Delta}(a) &= EU\left(S|t_S(a), a\right) - EU\left(D|t_D(a), a\right) \\
&= a\,u\left(t_S(a)\right) - t_S(a) - \sigma L^*\left(t_S(a)\right) \\
&\quad -a\,u\left(t_D(a)\right) + t_D(a) + \delta H\left(\alpha_D\right)\left(\beta\chi_C - \chi_N\right) \\
\text{So:} \quad \frac{\partial\bar{\Delta}}{\partial a} &= (a-\sigma)\,u'\left(t_S(a)\right)\frac{\partial t_S(a)}{\partial a} - \frac{\partial t_S(a)}{\partial a} + u\left(t_S(a)\right) \\
&\quad + \frac{\partial t_D(a)}{\partial a} - au'\left(t_D(a)\right)\frac{\partial t_D(a)}{\partial a} - u\left(t_D(a)\right) \\
&= u\left(t_S(a)\right) - u\left(t_D(a)\right) < 0
\end{aligned}$$

So D strictly dominates S for sufficiently large values of a. By symmetry, indifference point a_D is implicitly defined by:

$$\begin{aligned}
\lambda = a_D\,\left[u\left(t_S\left(a_D\right)\right) - u\left(t_D\left(a_D\right)\right)\right] + t_D\left(a_D\right) - t_S\left(a_D\right) \\
-\sigma L^*\left(t_S\left(a_D\right)\right) + \delta H\left(a_D\right)\left(\beta\chi_C - \chi_N\right) = 0
\end{aligned}$$

The equilibrium exists iff: $\overline{\Delta}\left(a_S\right) > 0$. By the definition of $\lambda = 0$:

$$\begin{aligned}
\bar{\Delta}\left(a_S\right) &= a_S\,\left[u\left(t_B\right) - u\left(t_D\left(a_S\right)\right)\right] + t_D\left(a_S\right) - t_B + \delta H\left(a_D\right)\left(\beta\chi_C - \chi_N\right) \\
&= a_D\,\left[u\left(t_D\left(a_D\right)\right) - u\left(t_S\left(a_D\right)\right)\right] + \sigma L^*\left(t_S\left(a_D\right)\right) + t_D\left(a_S\right) - t_B \\
&\quad + a_S\,\left[u\left(t_B\right) - u\left(t_D\left(a_S\right)\right)\right] + t_S\left(a_D\right) - t_D\left(a_D\right)
\end{aligned}$$

As shown below in the proof of Proposition 3.4, a_D grows larger as β increases. Note that $a_S\,\left[u\left(t_B\right) - u\left(t_D\left(a_S\right)\right)\right]$ is not a function of a_D. If the utility functions are such that $\frac{\partial}{\partial a}\left[t_D\left(a\right) - t_S\left(a\right)\right] \leq 0$, this is sufficient to show that $\overline{\Delta}\left(a_S\right) > 0$ for sufficiently large β. For example, this holds for $u(t) = \ln(t)$. This equilibrium exists even if $\beta = 1$ (i.e. there is no growth). Suppose $u(t) = \ln(t)$, $\delta = 0.9$, $A = 5$, $t_B = 3$, and $\sigma = 0.75$. Then R simulations (available on request from the author) demonstrate equilibrium existence in which $a_S = 3.75$ and $a_D \approx 4.96$. □

Continuation values

Let $t_E(a)$ denote equilibrium tariffs when the institution is in place. If the institution does not exist, then each state chooses $t_D(a)$ and $\tau_D(\alpha)$ in every time period. This yields continuation payoff:

$$\chi_N = \frac{1}{1-\delta}\int_1^A \left[(a-1)u\left(t_D(a)\right) - t_D(a)\right] dH(a)$$

The continuation payoff for home from the treaty being in effect is:

$$\chi_C = \int_1^A \left[au\left(t_E(a)\right) - t_E(a)\right] dH(a) - \sigma \int_{a_S}^{a_D} L^*\left(t_E(a)\right) dH(a) - \int_1^A u\left(\tau_E(\alpha)\right) dH(\alpha)$$
$$+ \sigma \int_{\alpha_S}^{\alpha_D} L^*\left(\tau_E(\alpha)\right) dH(\alpha) + \delta H\left(a_D\right)^2 \beta \chi_C + \delta\left[1 - H\left(a_D\right)^2\right] \chi_N$$
$$= \frac{\Psi}{1 - \delta\beta H\left(a_D\right)^2}$$

where $\Psi = \int_1^A \left[(a-1)\, u\left(t_E(a)\right) - t_E(a)\right] dH(a) + \delta\left[1 - H\left(a_D\right)^2\right] \chi_N$

Proof of Proposition 3.3. Recall that the binding is not violated if $a \leq a_S = \frac{1}{u'(t_B)} + \sigma$.

$$\frac{\partial a_S}{\partial t_B} = \frac{-u''\left(t_B\right)}{\left[u'\left(t_B\right)\right]^2} > 0 \quad \text{and} \quad \frac{\partial a_S}{\partial \sigma} = 1 > 0$$

□

Proof of Proposition 3.4. The institution is stable if $a < a_D$. By the implicit function theorem:

$$\frac{\partial a_D}{\partial t_B} = -\frac{\lambda_{t_B}}{\lambda_{a_D}} \quad \text{and} \quad \frac{\partial a_D}{\partial \sigma} = -\frac{\lambda_\sigma}{\lambda_{a_D}} \quad \text{and} \quad \frac{\partial a_D}{\partial \beta} = -\frac{\lambda_\beta}{\lambda_{a_D}}$$

As A grows larger, both $H\left(a_D\right)$ and $h\left(a_D\right)$ grow smaller. So for sufficiently large A:

$$\lambda_{a_D} = \left(a_D - \sigma\right) u'\left(t_S\left(a_D\right)\right) \frac{\partial t_S\left(a_D\right)}{\partial a_D} - \frac{\partial t_S\left(a_D\right)}{\partial a_D} - a_D u'\left(t_D\left(a_D\right)\right) \frac{\partial t_D\left(a_D\right)}{\partial a_D} + \frac{\partial t_D\left(a_D\right)}{\partial a_D}$$
$$+ u\left(t_S\left(a_D\right)\right) - u\left(t_D\left(a_D\right)\right) + \delta H\left(a_D\right) \beta \frac{\partial \chi_C}{\partial a_D} + \delta h\left(a_D\right)\left(\beta \chi_C - \chi_N\right)$$
$$= u\left(t_S\left(a_D\right)\right) - u\left(t_D\left(a_D\right)\right) + \delta H\left(a_D\right) \beta \frac{\partial \chi_C}{\partial a_D} + \delta h\left(a_D\right)\left(\beta \chi_C - \chi_N\right) < 0$$

$$\lambda_{t_B} = \sigma u'\left(t_B\right) + \delta H\left(a_D\right) \beta \frac{\partial \chi_C}{\partial t_B} > 0$$

$$\lambda_\sigma = \left(a_D - \sigma\right) u'\left(t_S\left(a_D\right)\right) \frac{\partial t_S\left(a_D\right)}{\partial \sigma} - \frac{\partial t_S\left(a_D\right)}{\partial \sigma} - L^*\left(t_S\left(a_D\right)\right) + \delta H\left(a_D\right) \beta \frac{\partial \chi_C}{\partial \sigma}$$
$$= -L^*\left(t_S\left(a_D\right)\right) + \delta H\left(a_D\right) \beta \frac{\partial \chi_C}{\partial \sigma} < 0$$

$$\lambda_\beta = \frac{\delta H\left(a_D\right) \Psi}{\left[1 - \delta\beta H\left(a_D\right)^2\right]^2} > 0$$

$$\text{So:} \quad \frac{\partial a_D}{\partial t_B} > 0 \quad \text{and} \quad \frac{\partial a_D}{\partial \sigma} < 0 \quad \text{and} \quad \frac{\partial a_D}{\partial \beta} > 0$$

□

Proof of Proposition 3.5. Recall that χ_C is the expected utility of a state from being a member of the cooperative regime. In equilibrium, for any pair (t_B, σ):

$$\chi_C = \frac{a_D\left[u\left(t_D\left(a_D\right)\right) - u\left(t_S\left(a_D\right)\right)\right] - t_D\left(a_D\right) + t_S\left(a_D\right) + \sigma L^*\left(t_S\left(a_D\right)\right)}{\delta H\left(a_D\right)\beta} + \frac{\chi_N}{\beta}$$

Let χ^* be the highest level that is achievable, given equilibrium play. Then an optimal pair (t_B, σ) solves $\chi_C\left(t_B, \sigma\right) = \chi^*$. The two first-order conditions for such an optimal pair (t_B, σ) are:

$$\frac{d\chi_C}{dt_B} = \frac{\partial \chi_C}{\partial t_B} + \frac{\partial \chi_C}{\partial a_D}\frac{\partial a_D}{\partial t_B} = 0$$
$$\frac{d\chi_C}{d\sigma} = \frac{\partial \chi_C}{\partial \sigma} + \frac{\partial \chi_C}{\partial a_D}\frac{\partial a_D}{\partial \sigma} = 0$$

This implies that:

$$\frac{\frac{\partial \chi_C}{\partial t_B}}{\frac{\partial \chi_C}{\partial \sigma}} = \frac{\frac{\partial a_D}{\partial t_B}}{\frac{\partial a_D}{\partial \sigma}}$$

So for any pair (t_B, σ) that generates $\chi_C\left(t_B, \sigma\right) = \chi^*$:

$$\frac{dt_B}{d\sigma} = \frac{-\left(\frac{d\chi_C}{d\sigma}\right)}{\frac{d\chi_C}{dt_B}} = \frac{-\left(\frac{\partial \chi_C}{\partial \sigma} + \frac{\partial \chi_C}{\partial a_D}\frac{\partial a_D}{\partial \sigma}\right)}{\frac{\partial \chi_C}{\partial t_B} + \frac{\partial \chi_C}{\partial a_D}\frac{\partial a_D}{\partial t_B}}$$
$$= \frac{-\frac{\partial \chi_C}{\partial \sigma}\left(\frac{\partial \chi_C}{\partial \sigma} + \frac{\partial \chi_C}{\partial a_D}\frac{\partial a_D}{\partial \sigma}\right)}{\frac{\partial \chi_C}{\partial t_B}\left(\frac{\partial \chi_C}{\partial \sigma} + \frac{\partial \chi_C}{\partial a_D}\frac{\partial a_D}{\partial \sigma}\right)} = \frac{-\frac{\partial \chi_C}{\partial \sigma}}{\frac{\partial \chi_C}{\partial t_B}}$$

where:

$$\frac{\partial \chi_C}{\partial \sigma} = \frac{1}{\delta H\left(a_D\right)\beta}\left[\frac{\partial t_S\left(a_D\right)}{\partial \sigma} - \left(a_D - \sigma\right)u'\left(t_S\left(a_D\right)\right)\frac{\partial t_S\left(a_D\right)}{\partial \sigma} + L^*\left(t_S\left(a_D\right)\right)\right]$$
$$= \frac{L^*\left(t_S\left(a_D\right)\right)}{\delta H\left(a_D\right)\beta} > 0$$
$$\frac{\partial \chi_C}{\partial t_B} = \frac{-\sigma u'\left(t_B\right)}{\delta H\left(a_D\right)\beta} < 0$$

$$\text{So:} \qquad \frac{dt_B}{d\sigma} = \frac{L^*\left(t_S\left(a_D\right)\right)}{\sigma u'\left(t_B\right)} > 0$$

□

Acknowledgements

Many thanks to B. Peter Rosendorff, who collaborated on an earlier version of this project. I also thank Renee Bowen, Ethan Bueno de Mesquita, Justin Fox, Raymond Hicks, Hyeran Jo, Stuart

Jordan, Soo Yeon Kim, Quan Li, Helen Milner, Jide Nzelibe, Krzysztof Pelc, Christina Schneider, Matthew Stephenson, and Jennifer Tobin. Previous versions of this paper were presented at the University of Chicago, Princeton University, Emory University, Stanford University's Graduate School of Business, and Texas A&M University. I am grateful for feedback from audiences at all of these presentations.

Funding

The author(s) received no financial support for the research and/or authorship of this article.

Notes

1. For ease of exposition, I focus here on tariffs. However, my model can easily be applied to non-tariff barriers.
2. For example, see Baccini (2010); Baccini et al. (2012); Haftel (2013); Hicks and Kim (2012); Kucik (2012).
3. See the codebook for Hicks and Kim (2012) for details of this example.
4. Because of this distinction between primary and secondary obligations, scholars sometimes refer to full compliance as 'first-order compliance', and partial compliance as 'second-order compliance' (Simmons, 1998).
5. Rosendorff (2005) assumes that the fine is imposed by a court that does not know the size of the violation. However, the fine is fixed from the perspective of a leader who must decide whether to defect.
6. In the Rosendorff and Milner (2001) and Rosendorff (2005) equilibria, the leader either meets the tariff binding or chooses a defection tariff. So the settlement tariff does not change in response to changes in the rigidity of the agreement.
7. Alternative accounts of flexibility argue that trade agreements create transparency about government policies for domestic audiences (Johns and Rosendorff, 2009; Mansfield et al., 2000; Milner et al., 2004). This in turn can reduce trade volatility, one of the key benefits of international trade agreements (Mansfield and Reinhardt, 2008). This mechanism is not inconsistent with my account. Since I do not directly model the behavior of voters, my model identifies the impact of rigidity on trade outcomes independent of effects generated by transparency.
8. I restrict attention to a two-player framework for both substantive and methodological reasons. First, despite the dominant role of the GATT/WTO in academic studies of international trade, the failure of trade negotiators to deepen trade concessions in the Doha Round of multilateral trade negotiations has led to a surge in preferential and regional trade agreements with restricted membership, such as NAFTA. Understanding the design of these agreements is important since they are the dominant mechanism for contemporary trade liberalization. Second, restricting the analysis to two players greatly simplifies the assumptions, exposition, and tractability of the model. This allows for a more nuanced understanding of the mechanisms that drive the analysis. It would be interesting to see how my results would extend to an n-player game. Increasing the number of players would ensure that the equilibrium cutpoint between settlement and defection (a_D) would change over time as a function of endogenous changes in the number of treaty members. Unfortunately, an n-player model is beyond the scope of this article.
9. Since this is a symmetric game, I assume that the two states have the same binding. This assumption does not affect the substantive results about the impact of depth and rigidity on full compliance and stability. However, it does affect the results on the optimal treaty design. It is trivial to show that if the two states are not required to commit to the same tariff binding, then each prefers a high binding (shallow concessions) for itself and a low binding (deep

concessions) for its partner. This means that the characteristics of an endogenous treaty will be affected by assumptions about the bargaining protocol that is used to negotiate the treaty.
10. This is akin to the separation of the bargaining and enforcement stages in Fearon (1998).
11. This interpretation is consistent with the model in Gilligan et al. (2010).
12. For equilibrium-as-institution arguments, see Calvert (1995) and Milgrom et al. (1990). See Rubinstein (1991) on models as understandings of strategic interactions.
13. Hicks and Kim (2012) code AD rules for a set of Asian trade agreements.
14. In their formal model of trade negotiations, Maggi and Staiger (2010: 5) show that it 'is never optimal to set damages high enough to make the exporter "whole"'.
15. There is certainly a cost to writing and joining a trade agreement. However, after a state has joined an agreement, this cost is sunk: it does not affect decisions about tariff levels and settlement because the cost cannot be recovered.
16. I suppress time subscripts for the sake of clarity.
17. Restrictions on the utility function are made explicit in the Appendix.
18. I assume that a and α are independently and identically distributed such that $a, \alpha \sim_{iid} U[1, A]$ for large A. The model results on tariff levels, full compliance, and stability (Lemma 3.2 and Propositions 3.1–3.4) are robust to asymmetric type distributions and tariff bindings. The details of this analysis are available upon request from the author.
19. Details about how growth is modeled can be found in the Appendix.
20. All results hold if the continuation payoff from defection is not a function of the value of the defection tariff. This includes tit-for-tat and temporary reversions to anarchy. Results are available from the author upon request.
21. Conditions on the existence of this equilibrium are specified in the Appendix.
22. This is usually referred to as a 'tariff overhang' (Busch and Pelc, 2010).
23. Of course, designing a trade agreement is a costly endeavor. We should not observe agreements in which these design costs outweigh the expected equilibrium trade benefits. A state has no incentive to write such a treaty. Including such a cost in the design stage does not affect the relationship between depth and rigidity in observable treaties (Proposition 3.5).
24. I thank Peter Rosendorff for extensive conversations about this result.
25. That is, I make no claim about what values of depth and rigidity a leader will choose. Rather, I am making claims about the relationship between these two attributes of treaty design.
26. My model does not allow me to assess the impact of the level of achievable benefits on full compliance and stability (the domestic pressure thresholds in Sections 3.3–3.4). Partial equilibrium reasoning would suggest that as states benefit more from a trade agreement, they will be more likely to comply with its terms and the agreement will be more stable. However, as states benefit more from trade cooperation, they should also design treaties that are deeper and more rigid. The overall effect of benefits from trade on compliance and stability are ambiguous because states can alter the terms of a treaty to affect final outcomes.
27. Of course, a leader does implicitly consider interest group preferences in my model because his beliefs about future domestic political pressure affect his expected utility from the trade agreement.
28. Von Stein (2005) explicitly argues that selection effects drive the empirical relationship. Simmons (2000) believes that the treaty has an independent effect of state behavior, but also argues that countries join Article VIII because they want to credibly signal their commitment to liberalization.
29. I thank Krzysztof Pelc for suggesting this example.
30. The scope of IMF conditionality has fluctuated over time. However, IMF agreements in the early 1990s reflected the prevalent economic belief (known as the Washington Consensus) that free market reforms were necessary for long-term growth in developing countries (Vreeland, 2008).

31. For challenges to Rose, see Goldstein et al. (2007), Mansfield and Reinhardt (2008), Subramanian and Wei (2007), and Tomz et al. (2007). For challenges to Von Stein, see Simmons (2000) and Simmons and Hopkins (2005).

References

Abbott KW and Snidal D (2000) Hard and soft law in international governance. *International Organization* 54: 421–456.

Alter KJ and Helfer LR (2010) Nature or nurture? Judicial lawmaking in the European court of justice and the Andean tribunal of justice. *International Organization* 64: 563–592.

Baccini L (2010) Explaining formation and design of EU trade agreements: the role of transparency and flexibility. *European Union Politics* 11: 195–217.

Baccini L, Dur A and Elsig M (2012) The politics of trade agreement design: the depth–flexibility trade-off. Working paper, London School of Economics.

Bown C (2012) Temporary trade barriers database. The World Bank. Available at: http://econ.worldbank.org/ttbd/.

Bown C and Tovar P (2011) Trade liberalization, antidumping, and safeguards: evidence from India's tariff reform. *Journal of Development Economics* 96: 115–125.

Busch ML and Pelc KJ (2010) Tariff bindings, tariff overhang, and trade remedies: policy flexibility at the WTO in tough political times. Working paper, Georgetown University.

Calvert RL (1995) Rational actors, equilibrium, and social institutions. In: Knight J and Sened I (eds) *Explaining Social Institutions*. Ann Arbor, MI: University of Michigan Press, pp.57–93.

Downs GW and Rocke DM (1995) *Optimal Imperfection? Domestic Uncertainty and Institutions in International Relations*. Princeton, NJ: Princeton University Press.

Downs GW, Rocke DM and Barsoom PN (1996) Is the good news about compliance good news about cooperation? *International Organization* 50: 379–406.

Fearon JD (1998) Bargaining, enforcement and international cooperation. *International Organization* 52: 269–306.

Fudenberg D and Tirole J (2000) *Game Theory*. Cambridge, MA: MIT Press.

Gilligan M (1997) *Empowering Exporters: Delegation, Reciprocity and Collective Action in Twentieth Century American Trade Policy*. Ann Arbor, MI: University of Michigan Press.

Gilligan M and Johns L (2012) Formal models of international institutions. *Annual Review of Political Science* 15: 221–243.

Gilligan M, Johns L and Rosendorff BP (2010) Strengthening international courts and the early settlement of disputes. *Journal of Conflict Resolution* 54: 5–38.

Goldberg P, Khandelwal A, Pavcnik N et al. (2010) Imported intermediate inputs and domestic product growth: evidence from India. *Quarterly Journal of Economics* 125: 1727–1767.

Goldstein J, Kahler M, Keohane RO et al. (2000) Introduction: legalization and world politics. *International Organization* 54: 385–399.

Goldstein J and Martin LL (2000) Legalization, trade liberalization, and domestic politics: a cautionary note. *International Organization* 54: 603–632.

Goldstein JL, Rivers D and Tomz M (2007) Institutions in international relations: understanding the effects of the GATT and the WTO on world trade. *International Organization* 61: 37–67.

Haftel YZ (2012) *Regional Economic Institutions and Conflict Mitigation: Design, Implementation, and the Promise of Peace*. Ann Arbor, MI: University of Michigan Press.

Haftel YZ (2013) Commerce and institutions: trade, scope, and the design of regional economic organizations. *Review of International Organizations* 8(3): 389–414.

Hasan R, Mitra D and Ramaswamy K (2007) Trade reforms, labor regulations, and labor-demand elasticities: empirical evidence from India. *Review of Economics and Statistics* 89: 466–481.

Hicks R and Kim SY (2012) Reciprocal trade agreements in Asia: credible commitment to trade liberalization or paper tigers? *Journal of West Asian Studies* 12: 1–29.

Johns L (2007) A servant of two masters: communication and the selection of international bureaucrats. *International Organization* 61: 245–275.

Johns L (2012) Courts as coordinators: endogenous enforcement and jurisdiction in international adjudication. *Journal of Conflict Resolution* 56: 257–289.

Johns L and Rosendorff BP (2009) Dispute settlement, compliance and domestic politics. In: Hartigan JC (ed) *Trade Disputes and the Dispute Settlement Understanding of the WTO: An Interdisciplinary Assessment*. Emerald Group Publishing, pp.139–163.

Krishna P and Mitra D (1998) Trade liberalization, market discipline and productivity growth: new evidence from India. *Journal of Development Economics* 56: 447–462.

Kucik J (2012) The domestic politics of institutional design: producer preferences over trade agreement rules. *Economics and Politics* 24: 95–118.

Kucik J and Reinhardt E (2008) Does flexibility promote cooperation? An application to the global trade regime. *International Organization* 62: 477–505.

Maggi G and Staiger RW (2010) Breach, remedies and settlement in trade agreements. Working paper, Yale University.

Mansfield ED, Milner HV and Rosendorff BP (2000) Free to trade: democracies, autocracies and international trade. *American Political Science Review* 94: 305–322.

Mansfield ED and Reinhardt E (2008) International institutions and the volatility of international trade. *International Organization* 62: 621–652.

Milgrom PR, North DC and Weingast BR (1990) The role of institutions in the revival of trade: the law merchant, private judges, and the champagne fairs. *Economics and Politics* 2: 1–23.

Milner HV, Mansfield ED and Rosendorff BP (2004) International trade and domestic politics: the domestic sources of international trade agreements and institutions. In: Benvenisti E and Hirsch M (eds) *The Impact of International Law on International Cooperation: Theoretical Perspectives*. Cambridge: Cambridge University Press.

Milner HV and Rosendorff BP (1996) Trade negotiation, information and domestic politics: the role of domestic groups. *Economics and Politics* 8: 145–189.

Rose AK (2004a) Do we really know that the WTO increases trade? *American Economic Review* 94: 98–114.

Rose AK (2004b) Do WTO members have more liberal trade policy? *Journal of International Economics* 63: 209–235.

Rose AK (2005) Does the WTO make trade more stable? *Open Economies Review* 16: 7–22.

Rosendorff BP (2005) Stability and rigidity: politics and the design of the WTO's dispute resolution procedure. *American Political Science Review* 99: 389–400.

Rosendorff BP and Milner HV (2001) The optimal design of international institutions: uncertainty and escape. *International Organization* 55: 829–857.

Rubinstein A (1991) Comments on the interpretation of game theory. *Econometrica* 59: 909–924.

Simmons BA (1998) Compliance with international agreements. *Annual Review of Political Science* 1: 75–93.

Simmons BA (2000) International law and state behavior: commitment and compliance in international monetary affairs. *American Political Science Review* 94: 819–835.

Simmons BA (2009) *Mobilizing for Human Rights: International Law in Domestic Politics*. Cambridge, New York: Cambridge University Press.

Simmons BA and Hopkins DJ (2005) The constraining power of international treaties: theory and methods. *American Political Science Review* 99: 623–631.

Smith JM (2000) The politics of dispute settlement design: explaining legalism in regional trade pacts. *International Organization* 54: 137–180.

Subramanian A and Wei SJ (2007) The WTO promotes trade, strongly but unevenly. *Journal of International Economics* 72: 151–175.

Tomz M, Goldstein JL and Rivers D (2007) Do we really know that the WTO increases trade? Comment. *American Economic Review* 97: 2005–2018.

Voeten E (2001) Outside options and the logic of security council action. *American Political Science Review* 95: 845–58.

Von Stein J (2005) Do treaties constrain or screen? Selection bias and treaty compliance. *American Political Science Review* 99: 611–622.

Vreeland J (2008) *The International Monetary Fund: Politics of Conditional Lending*. London: Routledge.