Depth versus Rigidity in the Design of International Trade Agreements*

Leslie Johns†

January 23, 2012

Abstract

I show that there is a trade-off between depth and rigidity in the design of international trade agreements. Increasing the depth of required cooperation both lowers the likelihood of compliance with the treaty and decreases the stability of the regime. In contrast, as the rigidity of an agreement increases (and there is less opportunity for tolerated defection from treaty obligations), the likelihood of compliance with the treaty increases and the stability of the regime decreases. Treaty designers must balance short-term compliance against the long-run stability of a cooperative regime. I argue that this implies a negative relationship between depth and rigidity in observed treaties. International trade agreements that mandate deep tariff reductions will be more likely to include flexibility mechanisms—such as escape clauses and antidumping procedures—than agreements that require shallow cooperation.

*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, Jide Nzelibe, Christina Schneider, Matthew Stephenson, and Jennifer Tobin. Previous versions of this paper were presented at the University of Chicago, Princeton Unversity, 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.

†Department of Political Science, UCLA. 4289 Bunche Hall. Los Angeles, CA 90095.

States create international agreements in order to secure mutual gains from cooperation. Many examples of international cooperation—including international trade, military alliances, and arms reduction treaties—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 her ability to protect import-competing industries. However, a system of tariff reductions increases overall social welfare by benefiting consumers and exporting industries. 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 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 nontariff barriers, antidumping duties, safeguards protections, etc.—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 her international trade obligations and protect domestic industries. Similarly, contributions to a military alliance benefits 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.

This means that any international agreement must have two basic components. First, states must negotiate the *depth* of cooperation by specifiying what specifically constitutes cooperation. Second, states must decide how to treat states 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 both 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 key outcomes when designing treaties. First, the design of a treaty affects the level of state *compliance*—whether the actual behavior of the state conforms to

behavior prescribed by the cooperative treaty.[1] Second, depth and rigidity affect the *stability* of the cooperative regime—the ability of the regime to endure over time.

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

I show that increasing the depth of cooperation (by choosing lower tariff bindings) both lowers the likelihood of compliance with the treaty in a given period and decreases the long-term stability of the regime. In contrast, as the rigidity of an agreement increases (and there is less opportunity for tolerated defection from the terms of the treaty), the likelihood of compliance with the treaty increases and the stability of the regime decreases. Changes to the depth or rigidity of the treaty have indeterminate effects on the overall benefits that are generated by the trading regime for member governments. However, increasing either factor decreases the stability of the regime. I show that these fundamental trade-offs in compliance and stability affect the rational design of international agreements. There will be a negative correlation between depth and rigidity in treaties that are actually signed by states—and hence observed by researchers—when we control for the total level of cooperative benefits that are achievable by treaty members. That is, I explicitly account for the endogenous design of international agreements and show that agreements that require deep levels of cooperation will be more likely to include flexibility mechanisms—such as escape clauses and antidumping procedures—while agreements that require shallow cooperation will be more rigid.

[1] I focus here on first-order compliance—whether a treaty member abides by the tariff bindings specified in the treaty—rather than on second-order compliance. See Simmons (1998*a*) on the distinction between these two concepts.

1 Depth and Rigidity in International Trade Agreements

Recent studies of the political economy of international trade have emphasized the role of treaty design on political and economic outcomes. One key element of treaty design is the *rigidity* of the agreement—the degree to which defections from prescribed behavior are tolerated. Traditional IR scholars argue that strong enforcement mechanisms are necessary to support deep levels of international cooperation. Since states face short-term incentives to defect from cooperative behavior, large punishments for noncooperative behavior are believed to be necessary to deter defection.[2] However, recent scholarship has challenged this view by arguing that flexible, less rigid agreements will lead to more cooperation in the long-run.[3]

The study of international agreements cannot be divorced from domestic politics.[4] Sometimes a country will experience tough times and a political leader will face strong pressure to violate trade obligations in order to protect domestic industries. If a state has joined an agreement and the domestic political conditions shift, the state will feel pressure to rescind the earlier concessions. Under a rigid agreement, leaders must choose between compliance or exit from the cooperative regime. Rigid enforcement regimes leave a leader with little policy discretion during tough times. In order to deal with this problem, many agreements introduce flexibility-enhancing devices, such as escape clauses, exceptions, relief under special or unforeseen circumstances, safeguard provisions, and antidumping procedures. These flexibility-enhancing devices allow states to sign agreements and then temporarily violate the strict requirements of the treaty (at a cost) during periods of heightened political pressure.[5] Defection is temporarily tolerated, although states are expected to return to cooperative behavior once a crisis has passed. Under rigid regimes, a leader's only options are to comply or to exit. Flexibility mechanisms reduce this rigidity by giving leaders the option of commiting temporary violations while remaining a member of the regime. The knowledge that temporary noncompliance is possible in later periods–even at some cost—makes a state more willing to sign such an agreement.

[2]For example, see Downs, Rocke and Barsoom (1996).
[3]For example, see Rosendorff and Milner (2001).
[4]Goldstein and Martin (2000).
[5]Downs and Rocke (1995, 76-104).

Empirical scholars have demonstrated that preferential trade agreements vary tremendously in their level of rigidity.[6] Consider the specific example of safeguard provisions, which allow a treaty member to violate its tariff concessions if an surge in imports harms or threatens a domestic industry. First, trade agreements vary in the criteria used to identify such import surges. Some treaties allow a single member to unilaterally invoke provisions, while others require bilateral consultations or even specify objective standards that are monitored by a treaty body. Second, trade agreements vary in the discretion they grant in responding to import surges. Some treaties impose time-limits and/or explicitly require compensation to affected trade partners, while others impose no constraints whatsoever.[7] Similar variation exists in the design of other flexibility-enhancing devices, such as antidumping and countervailing duties.

In contrast to earlier theoretical work, I treat the rigidity of an agreement as a continuous measure. Suppose that temporary noncompliance followed by a return to future cooperation comes with a cost. If the agreement is extremely rigid, then the costs of temporary tolerated violation are so excessively large that a leader will always prefer to exit a cooperative regime when he commits a violation. In contrast, if the agreement is very flexible, then there is little cost to remaining a member of the regime after temporarily violating a treaty. The overall rigidity of an agreement can be characterized by how costly it is to "purchase forgiveness" from other treaty members when a temporary violation occurs. I refer this process as the "settlement" of the trade dispute that is created by the temporary violation. Rigid treaties will require large settlements for temporary noncompliance, and more flexible agreements will require smaller settlements. While flexibility mechanisms—such as escape clauses, safeguards, antidumping procedures, and other exceptions—lower the likelihood and level of compliance, they also enhance the stability of trading regimes by decreasing the rigidity of the agreement.[8] Rather than exiting the treaty regime during tough times, a leader will violate, voluntarily settle the dispute in subsequent negotiations, and then return to cooperative behavior in the future.[9]

[6] For example, see Haftel (2010), Hicks and Kim (2010), and Kucik (2011).

[7] See the codebook for Hicks and Kim (2010) for details of this example.

[8] See Rosendorff and Milner (2001) and Rosendorff (2005).

[9] Alternative accounts of the value of flexibility provisions argue that trade agreements create transparency about government policies for domestic audiences (Johns and Rosendorff 2009; Mansfield, Milner and Rosendorff 2000; Milner, Mansfield and Rosendorff 2004). This in turn can reduce trade volatility, one of the key benefits of international

A second key element of treaty design is the *depth* of cooperation—the level of tariff concessions in a trade agreement. It has already been well-established that there is a tradeoff between the depth of international agreements and the level of compliance.[10] As an agreement requires deeper levels of cooperation from states—i.e. requires more significant changes in behavior from what the state would do in the absence of the agreement—there will be less compliance with its provisions, *ceteris paribus*. However, the relationship between depth of cooperation and the stability of an institution has been unexplored. Are agreements that require deeper levels of cooperation more or less stable than agreements with shallower demands on state behavior? I find that deeper cooperation leads to less institutional stability because of the mechanism discussed above. By making more onerous demands on state behavior, a deep treaty makes exit from the regime more desirable when a leader faces tough times.

Obviously states will still find it beneficial to design treaties with some degree of depth and rigidity. After all, the most stable regime possible is one that demands no changes in the behavior of states; such a regime also yields no cooperative benefit. Depth and rigidity in international agreements can be beneficial in promoting cooperation. However, the designers of treaty regimes must balance short-term compliance against the long-term stability of the institution. Treaties are endogenous. As scholars, we can only observe the agreements that states find it beneficial to write and sign. The design of international law must take account of the compliance-stability tradeoff.

I argue that if we can control for the total level of cooperative benefits that are achievable by treaty members, then observed treaties–those trade agreements that are actually written and signed by states–will have a negative relationship between these two factors. Deep tariff concessions will be accompanied by flexibility provisions, while more rigid agreements will have shallower levels of trade concessions. This relationship is supported by qualitative accounts of the development of the Caribbean Community and the 2011 EU-South Korea trade agreement.[11] After presenting my theoretical model, I discuss some recent quantitative studies that provide additional empirical

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.

[10] Downs, Rocke and Barsoom (1996).

[11] See Baccini et al. (2011).

support for my model.

2 Theory

I consider the strategic behavior of two states, which I call "home" and "foreign."[12] The game has two stages. In the first stage, the states bargain over the text of a cooperative treaty. To fully specify the treaty they must specify both the depth of cooperation and the rigidity of the treaty in addressing violations. Depth of cooperation is represented by the level of the tariff binding, t_B.[13] 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 the tariff bindings, 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 overall 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 in a given period it must also choose how much (if any) compensation it will pay to its trading partner. This subgame is infinitely-repeated.[14]

[12]The choice to restrict attention to a two-player framework is done 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.

[13]Since this is a symmetric game, I assume that the two states must commit to the same tariff level. This assumption does not affect any of the substantive results about the impact of depth and rigidity on compliance and stability. However, it does affect the results on negotiations over 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 specific assumptions about the bargaining protocol that is used to negotiate the treaty.

[14]This is akin to the separation of the bargaining and enforcement stages in Fearon (1998).

2.1 Settlement of Trade Disputes

Before proceeding to a description of the cooperation game, I think it necessary to briefly elaborate on my reduced-form model of the settlement of trade disputes. Treaties vary significantly in their dispute settlement procedures. Some trade agreements contain no formal procedures whatsoever, while others—such as the WTO and NAFTA—create specialized judicial institutions. However, a common element in all aspects of international trade law is an emphasis on the peaceful settlement of disputes. Differences between states may be resolved before a trial occurs, during litigation, or even after a judicial institution has made a ruling. In previous work I explicitly modelled the process by which states negotiate settlements before and after legal rulings.[15] Here I construct a simple measure that allows me to separate the effect of the rigidity of the treaty from variation in the size of the dispute—as proxied by the magnitude of losses from the treaty violation.

Most trade treaties explicitly require disputants to conduct negotiations prior to using formal dispute settlement procedures. For instance, the ASEAN agreement requires that each member "accord adequate opportunity for consultations regarding ... any matter affecting the implementation of this agreement."[16] The member that has been approached "shall give full due considerations to the proposals made to it."[17] If no satisfactory agreement is found in consultations, the ASEAN dispute panel must try to facilitate a settlement. Only after this process has occurred can a member temporarily suspend concessions in order to encourage the other side to settle the dispute. The ASEAN-China agreement is even more explicit in requiring negotiations. It requires that if "the party complained against fails to bring the measure ... into compliance ... that party shall, if so requested, enter into negotiations with the complaining party with a view to reaching a mutually satisfactory agreement on any necessary compensatory adjustment."[18] If no agreement on compensation is reached via negotiations, then "the complaining party may request the original arbitral tribunal to determine the appropriate level of any suspension of concessions or benefits."[19]

[15] See Gilligan, Johns and Rosendorff (2010) and Johns (2012).

[16] Agreement on the Common Effective Preferential Tariff (CEPT) Scheme for the ASEAN Free Trade Area (AFTA), Article 8(1).

[17] Agreement on the Common Effective Preferential Tariff (CEPT) Scheme for the ASEAN Free Trade Area (AFTA), Article 8(2).

[18] Agreement on Dispute Settlement Mechanism of the Framework Agreement on Comprehensive Economic Co-Operation Between the Association of South East Asian Nations and the People's Republic of China, Article 13(2).

[19] Agreement on Dispute Settlement Mechanism of the Framework Agreement on Comprehensive Economic Co-

Settlements must be mutually acceptable and can take the form of adjustments to the trade measure, technical assistance to reduce the effects of the offending measure, and even withdrawal of equivalent concessions. The losses incurred by the complaining state are mitigated by some form of settlement either before or after adjudication, or—as a last resort—via a carefully calibrated tariff retaliation.[20]

I interpret the general requirement that negotiated settlements are to be sought as follows. The offending measure induces a loss to the complaining state, and a gain to the offending state. A settlement is an outcome in which this loss is mitigated and some of the gains are reduced. In the rare instance in which the complaining state withdraws previously granted concessions, the effect is to reduce the losses incurred, and to apply a loss to the offending state. Settlement therefore has the effect of redistributing gains and losses that were generated by the measure in the first place.

Ssettlement is a political process. Trade agreements can provide guidelines throughout this process, but rarely prescribe precise levels of compensation. Nonetheless, we can conceptualize the severity of expected settlement outcomes as an attribute of the treaty regime. If the treaty is very flexible, then states will have a permissive attitude toward temporary violations of trade concessions. In such a trade regime, states can violate their obligations during tough times and return to future cooperation with little compensation to injured parties. Such a regime is characterized by a low value of σ. In contrast, very rigid treaties will not be tolerant of violations. Treaty violators will suffer dearly in dispute negotiations and make large settlements (high σ) before cooperation can resume.

However, I assume that trading partners will never be "made whole" following a trade violation; i.e. $\sigma < 1$. This assumption is made 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 fundamental structure of prisoners' dilemma payoffs is preserved and my results are

Operation Between the Association of South East Asian Nations and the People's Republic of China, Article 13(3).

[20]This approach to dealing with disputes mirrors the approach taken at the WTO. Schwartz and Sykes (2002) argue persuasively that in the WTO there is little doubt that in the case of a violation, the countries are required to attempt to negotiate a settlement and pay compensation if necessary. Jackson (2004) has taken issue with this view and insists that there is no right to buy out or compensate a member in exchange for a violation. Nonetheless, he admits that the system explicitly has as its objective a "settlement" of the dispute, but such a settlement must be in conformity with the rulings of the DSU and with the text of the agreement.

not being driven by opportunities for efficient breach within the system. Even if a large settlement is reached, a state is still a "sucker" it if cooperates while its partner does not.[21] This assumption is also desirable for empirical reasons. The legal standard of "equivalent concessions" that is embodied in many treaty texts means that injured states cannot secure more than the losses they have actually 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 not unreasonable to think that any settlement will be a fraction of the damages incurred.

It matters little for my purposes when the states reach a dispute settlement. This can occur before, during, or after adjudication.[22] In all cases settlement will be some fraction of the loss that the plaintiff state has incurred. The magnitude of this fraction is to a large degree a parameter of the agreement itself. Rigid agreements are intolerant of defection and require large settlements or quick withdrawals of the offending measures. Flexible agreements will permit smaller settlements and perhaps longer periods of adjustment in order to return to compliance.

Finally, it is worth emphasizing that there are many different ways a reader can interpret the settlement parameter σ. For those who adopt a literal attitude towards the use of game theoretic models in the study of institutions, the settlement parameter can represent payoffs of an unmodelled subgame in which states actively negotiate and litigate under the terms of a treaty text.[23] A more adventurous reader can adopt an equilibrium-as-institution perspective and interpret σ as representing a set of shared understandings about what constitutes appropriate behavior.[24] Either interpretation is compatible with my arguments and findings.

2.2 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

[21] In their formal model of trade negotiations, Maggi and Staiger (2010, 4) show that it "is never optimal to set damages high enough the make the exporter 'whole'."

[22] Busch and Reinhardt (2000) report that two-thirds of WTO disputes are settled before a final ruling of the Dispute Settlement Body.

[23] This interpretation is consistent with the model in Gilligan, Johns and Rosendorff (2010). Details of the equivalence are made explicit in Johns (2011).

[24] For equilibrium-as-institution arguments, see Calvert (1995) and Milgrom, North and Weingast (1990). See Rubinstein (1991) more generally on models as representations of *understandings* of strategic interactions.

choose tariff levels. In each period, let t denote the tariff chosen by the home government, and τ denote the tariff chosen by the foreign government.[25] Each government cares about the welfare of its import-competing firms, exporting firms, 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 benefits to an import-competing firm of the home tariff, t, is an increasing concave function, $u(t)$.[26] So import-competing firms always prefer higher tariffs on foreign goods, but their marginal benefit from a tariff declines as the tariff grows larger. Exporting firms care about the tariff in the foreign country, 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. I let a denote the political pressure experienced by the home government, and α denote the political pressure experienced by the foreign government.[27] So if the home government experiences high political pressure—i.e. 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 level of pressure each might face in any future period. So the one-period utility functions of the home and foreign government—W and W^*, respectively—are as follows:

[25] I suppress time subscripts for the sake of clarity.

[26] Restrictions on the form of the utility function are made explicit in the Appendix.

[27] I assume that a and α are independently and identically distributed such that $a, \alpha \sim_{iid} U[1, A]$ for large A.

$$
\begin{aligned}
W(t,\tau,a) &= a\,u(t) - t - u(\tau) \\
W^*(t,\tau,\alpha) &= \alpha\,u(\tau) - \tau - u(t)
\end{aligned}
$$

After observing its own current level of political pressure, each government simultaneously chooses a tariff.[28]

If the foreign country violates the trade agreement by choosing a tariff that is too high (i.e. $\tau > t_B$), then the home government's losses are: $L(\tau) = W\left(t, t_B, a\right) - W(t,\tau,a) = u(\tau) - u\left(\tau_B\right)$. Under the terms of the treaty, the foreign state must pay $\sigma L(\tau)$ to the home country for such a violation. Similarly, if the home country violates the agreement by choosing $t > t_B$, then the foreign state's losses are: $L^*(t) = W^*\left(t_B, \tau, \alpha\right) - W^*(t,\tau,\alpha) = u(t) - u\left(t_B\right)$. The home country must pay $\sigma L^*(t)$ to the foreign country to abide by the treaty terms. Decisions about whether to settle—by paying the required level of compensation—are made voluntarily by each state.[29] A treaty member can choose to both violate the tariff binding and to 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 over time of pro-compliance domestic constituencies.[30] While governments always face some pressure to protect import-competing industries, we might still believe that the benefits of trade cooperation can grow over time because of expansions in the volume of trade, reductions in the cost of transportation, or long-term changes in the allocation of resources across sectors.[31]

[28] This form of the government's objective function is isomorphic to models derived from profit- and utility-maximizing firms and individuals engaged in market interactions in each country. For instance, see Bagwell and Staiger (1999). The weighting of the interests of the import-competing sector in government's returns is consistent with both the "politically optimal objective function" of Baldwin (1987) and the derived political support functions from models with campaign contributions such as Grossman and Helpman (1994). I ignore the tariff revenue by assuming that it is simply redistributed back to all individuals uniformly.

[29] The decision about whether to pay the required settlement is thus a decision about second-order compliance (Simmons 1998*b*).

[30] For example, see Alter and Helfer (2010), Dai (2005), Gilligan (1997), and Simmons (2009).

[31] Details about how this is growth is modelled are included in the Appendix.

2.3 Equilibrium Behavior in the Cooperation Subgame

In infinitely-repeated games, most equilibrium concepts allow players to condition their actions on all components of the history of play. The well-known "folk theorems" of infinitely-repeated games have established that such a framework can support a large diversity of equilibrium behavior.[32] For example, there exists an equilibrium of my model in which each government always chooses the tariff that maximizes its one-period utility under political pressure, which I call the "defection tariff." In this equilibrium, countries never pay back any of the losses incurred by the trading partner and government behavior is never constrained by the tariff binding. I refer to this as an anarchic equilibrium because the treaty has no meaningful effect on state behavior.

However, the interesting theoretical question is how the design of the cooperative regime can promote cooperation between states. The ability of states to cooperate will depend on their patience. The more patient a state is, the more willing it is to sacrifice the short-term benefit of a high unilateral tariff for the long-term benefits of joint cooperation. I consider an equilibrium in which each state conditions its behavior on whether past violations have occurred and whether settlement occurred. In order to examine the limits on institutional performance, I assume grim trigger punishment. Namely, I assume that if one state violates the tariff binding and refuses to pay the required share, σ, of its partner's loss, then the cooperative tariff regime collapses and states choose defection tariffs in all future periods. The collapse of the treaty is equivalent to a return to the anarchic equilibrium.

This means that I examine equilibria in which three different action profiles are possible when there has been a past history of cooperation. First, a state can comply (action C) by choosing a tariff that does not exceed the tariff binding ($t \leq t_B$ and $\tau \leq \tau_B$). Second, a state can exceed the binding but then settle (action S) by repaying some of the loss mandated by the treaty. Either of these first two actions ensures that the treaty will remain in effect for the next period. Finally, a state can defect (action D) by violating the tariff binding and refusing to pay back any of the loss. This defection is punished by the collapse of the cooperative regime (the grim trigger punishment) for all future periods. Trade still occurs in future periods, but each government always engages in anarchic

[32] Fudenberg and Tirole (2000, 150-160).

behavior by choosing the tariff that maximizes its one-period utility under political pressure. So I do not assume that states will always comply with the terms of the treaty by choosing low tariffs and then settling if a violation occurs. Rather, my treaty is self-enforcing in the sense that any settlement that is paid is chosen voluntarily.

In the Appendix I fully characterise the Bayesian Nash equilibrium of this tariff setting game for a given treaty design. Each government policymaker will choose whether to comply (C), violate and settle (S), or violate and exit (D). This decision will be affected by the level of political pressure in the given 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 1. *There exists a Bayesian Nash equilibrium in which low types ($a < a_S$) comply (C), moderate types ($a_S \leq a \leq a_D$) violate the tariff binding and then settle (S), and high types ($a_D < a$) defect and leave the institution (D).*[33]

Rosendorff (2005) studies a similar problem and generates equilibrium behavior that is not dissimilar. However, the emphasis in that paper was to compare two possible institutional structures: trade agreements with or without dispute settlement procedures. He finds that agreements with such procedures are more stable, so flexibility enhances stability.[34] My emphasis here is quite different. First, I am interested in how both compliance and stability are affected by both rigidity and depth. Second, I examine how these relationships affect the characteristics of treaties that are designed endogenously by rational states.

The first implication of my model relates the rigidity of the treaty to the severity of violations chosen by the governments in equilibrium.

Lemma 1. *If a state decides to violate the trade agreement and not pay a settlement (D), then its tariff is increasing in the size of political pressure from import-competing industries. If a state decides to violate the agreement and pay settlement (S), then its tariff is increasing in the size of political pressure from import-competing industries and decreasing in the rigidity of the agreement (σ).*

[33] Conditions on the existence of this equilibrium are specified in the Appendix.

[34] See also Rosendorff and Milner (2001). In Proposition 3 below, I generalize this finding by exploring how the degree of stability varies with the degree of flexibility.

A more rigid trade agreement means that if a state decides to violate the treaty and offer a settlement, then the severity of the violation (the amount by which the chosen tariff exceeds the mandated binding) will be lower than if the treaty were more flexible.

Figure 4 shows the equilibrium graphically. On the horizontal axis is the level of political pressure experienced by the home country, a. The vertical axis is the tariff level. The horizontal line at t_B indicates the tariff binding that is mandated by the trade agreement. Suppose that the government faces low levels of political pressure. If the government were to choose its optimal tariff without regard to its international obligations, then the tariff that it would choose, t_D, is lower than the treaty binding.[35] When the government faces little to no pressure from import-competing industries, its trade policy is unconstrained by the treaty and the state complies with the binding. However, as political pressure increases, the government faces an enhanced incentive to increase its tariff on the imported good. For sufficiently high levels of political pressure the government's unconstrained ideal tariff (t_D) breaches the binding. In order to stay in compliance with the treaty, the government will need to restrict its tariff to level t_B. This is indicated by the flat portion of the bold line. As political pressure grows even larger, the political benefits of applying a tariff above t_B and then negotiating a settlement with the foreign government outweigh the benefits from full compliance. The optimal tariff in these circumstances, denoted by t_S, rises with the level of political pressure, but always lies below the pure defection tariff (i.e. t_D, the tariff that the government would choose in the absence of a treaty). Finally, if political pressure becomes very large, exit from the treaty becomes the optimal action. The government will choose the pure defection tariff t_D without settling the dispute.

[Insert Figure 1 here.]

2.4 Compliance and Stability in the Cooperation Subgame

Before I explore the negotiations over the parameters of the agreement, I need to know how changing those attributes affect the behavior of the signatory states. More specifically, I am interested in the effect of changes in the depth of bindings and the rigidity of the agreement on both the likelihood of

[35] This is usually referred to as a "tariff overhang" (Busch and Pelc 2010).

full compliance (C) and exit from the treaty regime (D). Operationally, this involves investigating the effects of the depth and rigidity variables on the equilibrium behavior thresholds a_S and a_D. A rise in threshold a_S means that the likelihood of full compliance increases. A government is more likely to choose tariffs at or below its binding level as a_S increases. A rise in threshold a_D means that a government is less likely to exit from the agreement. The trading regime is more stable as a_D increases.

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

Proposition 2. *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 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 periods. This makes each state less likely to violate in the first place. However, a lower tariff binding requires a deeper level of cooperation from treaty members. Compliance in general 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. The level of both depth and rigidity affects the threshold a_D, which determines the likelihood that the cooperative trading regime collapses and states return to anarchic behavior.

Proposition 3. *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 exiting the regime and foregoing 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 states are more willing to forgo the future benefits of trade cooperation by defecting today and refusing to pay a settlement (D). The set of political pressure values that make a state exit the regime 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, states that face large pressure to violate the tariff binding will find it more beneficial to exit the cooperative regime than to buy forgiveness by settling. Leaving the system becomes more attractive as the rigidity of the treaty increases. Both depth and rigidity reduce the stability of the agreement.

The results from both Propositions 2 and 3 are demonstrated graphically in Figures 2 and 3. In Figure 2, 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 levels of cooperation are required by the treaty. The solid line denotes equilibrium tariffs under the original binding t_B. The dashed line represents equilibrium tariffs when the binding is lowered to $\overline{t_B}$. As shown in Proposition 2, the lower threshold decreases from a_S to $\overline{a_S}$, which ensures that the region of full compliance shrinks from C to $\bar{C}$. 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. Increasing the depth of the treaty increases the size of defect region and makes exit from the regime more likely. This reduces the stability of the agreement. Trade agreements that demand deeper levels of cooperation get less compliance and have a higher risk of collapse, *ceteris paribus*.

[Insert Figure 2 here.]

In Figure 3, I explore the impact of increasing the rigidity of the trade agreement from σ to $\overline{\sigma}$. The solid line denotes equilibrium tariffs under the original level of rigidity, σ. The dashed line represents equilibrium tariffs when rigidity is increased to $\overline{\sigma}$. Optimal settlement tariffs are reduced from t_S to $\overline{t_S}$. This follows naturally: if a state is going to be more severely penalized for a given violation, then it will violate less by choosing a lower tariff. Increased rigidity affects the likelihood of full compliance and stability by changing the two threshold values. First, the zone of full compliance increases in size from C to $\bar{C}$ because the lower threshold a_S rises to $\overline{a_S}$.

Second, the upper threshold shifts down from a_D to $\overline{a_D}$ as rigidity increases. This increases the region in which states exit the trading regime. So enhanced rigidity reduces the stability of the regime. Additionally, the central region—the levels of political pressure for which there is tolerated violation and settlement (S)—shrinks in size, which reduces the likelihood of partial compliance. So increasing the rigidity of the treaty increases the likelihood of both full cooperation and collapse of the cooperative regime.

[Insert Figure 3 here.]

2.5 Depth versus Rigidity in Bargaining Outcomes

Changes in the design of the trade agreement clearly affect both compliance and the stability of the cooperative regime. What types of treaties will states negotiate prior to engaging in the cooperation subgame? I assume that when negotiating a treaty in the first stage of the overall game, states do not know the precise levels of political pressure that they will face in future periods. Each leader knows that he will need to satisfy exporters, import-cometing industries, and consumers. However, as in the basic stage game, he knows that pressure from each of these groups will fluctuate in response to political and economic shocks. The treaty is designed behind a Rawlsian "veil of ignorance" in which the leader knows that future shocks can occur but does not know the precise level of these shocks.

This means that the home and foreign state are identical in the first stage. They have the same preferences over the design of the treaty. An optimal treaty will maximize each state's expected payoff from being a member of a treaty regime. This means that any bargaining game can be reduced to a decision-theoretic framework in which a leader maximizes her utility from being a member of the cooperative treaty by choosing the depth and rigidity of the treaty. This utility is affected by both the short-term benefits from compliance, and the ability of the regime to provide long-term benefits by remaining stable over time.

The model structure that is established above is sufficiently general that it is not possible to explicitly characterize the optimal levels of depth and rigidity. However, when a state is writing an optimal treaty, there will be a limit to the level of cooperative benefits that is achievable. This level

will be affected by many factors, including the patience of the players, the opportunities for growth over time, and the shape of the utility function. Nevertheless, I show that in any optimization problem, there will be a negative relationship between depth and rigidity in optimal treaties.[36] That is, in order to maintain a given level of cooperative benefits, an increase in one parameter must be accompanied by a decrease in the other parameter. This leads to the following formal result.

Proposition 4. *Controlling for the total level of cooperative benefits that are achievable by treaty members, there will be a negative relationship between depth and rigidity when states design optimal trade agreements—increased depth must be accompanied by decreased rigidity, and vice versa.*

The optimal levels of depth and rigidity can vary across agreements.[37] I have no reason to think that the observed balance between compliance and stability is necessarily the same across all trade agreements since different groups of states are likely to vary in terms of their level of achievable cooperative benefits. However, my theory implies that if we control for the level of cooperative benefits, then there will be a negative relationship between depth and rigidity in the set of observable treaties. Regardless of how states decide to balance short- versus long-term benefits in a given situation, deeper concessions must be accompanied by increased flexibility. So an empirical implication of my theory is that the levels of depth and rigidity in observed treaties should be negatively related when appropriate variables control for the "achievability" of cooperative benefits.

The assumption that leaders design the treaty behind a Rawlsian "veil of ignorance" greatly simplifies the analysis of treaty negotiations. Skeptics might argue that this simplifying assumption overlooks the important role that domestic interest groups can play in shaping the design of treaties. A more complex model of treaty negotiations might allow exporters and import-competing industries to lobby their home governments over the design of the trade agreement. Kucik (2011) argues that exporters—who want open markets abroad—will prefer deep concessions and rigid agreements, while import-competing industries—who want closed markets at home—will prefer shallow conces-

[36] Many thanks to B. Peter Rosendorff and seminar participants at Stanford's Graduate School of Business for extensive conversations about this result.

[37] That is, I make no claim regarding what observed values of $\left(t^B, \sigma\right)$ will be. Rather, I am making claims about the relationship between these two elements of treaty design.

sions and flexible agreements. Various political factors—such as industry concentration, resources, and political mobilization—will affect the ability of each group to secure its preferred outcome. Nonetheless, the need to satisfy these competing domestic interest groups suggests that leaders probably be forced to make a trade-off: deep concessions will need to be balanced against flexibility to maintain political support from both exporters and import-competing industries.

3 Empirical Support for the Theoretical Model

Recall that we examined the impact of two design variables—depth and rigidity—on two outcomes—compliance and stability. Additionally, Proposition 4 suggests an important relationship between the two treaty design elements. Systematic empirical tests of my theory are beyond the scope of this paper. Nonetheless, previous empirical studies support many aspects of my theory.

Kucik and Reinhardt (2008) provides strong support for the relationship between rigidity and stability. This paper is an empirical test of the theoretical arguments in Rosendorff (2005). Kucik and Reinhardt examine whether the presence of antidumping (AD) procedures—domestic rules that sometimes allow states to violate their trading commitments without abrogating the treaty itself—affects membership in a trade regime. They argue that the presence of domestic AD procedures significantly increases the likelihood that a state will join and remain a member of the WTO. Additionally, they find that new members of the WTO are more likely to adopt antidumping procedures than non-members. Kucik and Reinhardt do not measure variation within AD procedures; i.e. how "flexible" the rules are. Nevertheless, their analysis provides strong support for the claim in Proposition 3 that rigidity decreases the stability of trading regimes.

There is also some support in existing empirical studies for Proposition 4, which posits a negative relationship between rigidity and depth in observable trade agreements. Kucik and Reinhardt (2008) find that AD rules—which reduce rigidity—make states more likely to accept and implement lower tariff rates, which implies deeper levels of cooperation. Similarly, Smith (2000) finds that trade agreements that require deeper levels of integration are more likely to include flexibility-enhancing devices, such as third-party review of state disputes, and Haftel (2010) shows that in a broad set of 25 regional economic organizations over two decades, deeper levels of cooperation are associated

with decreased rigidity.[38] Finally, Baccini et al. (2011) develop sophisticated techniques to empirically model the endogenous relationship between treaty design elements. They find strong support for the claim that there is a tradeoff between depth and rigidity in observable treaties.

To my knowledge, we lack systematic evidence on the relationship between rigidity and compliance in trade agreements. Additionally, it remains to be seen whether there is empirical evidence to support the theoretical relationships between depth and the two outcome variables—compliance and stability. Nevertheless, previous empirical studies clearly provide support for other elements of my theoretical framework.

4 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. Increasing the depth of cooperation both lowers the likelihood of compliance with the treaty and decreases the stability of the regime. In contrast, as the rigidity of an agreement increases (and there is less opportunity for tolerated defection from the terms of the treaty), the likelihood of compliance with the treaty increases while the stability of the regime decreases. Since states must balance compliance and stability when designing cooperative regimes, I argue that there will be a negative relationship between depth and rigidity in observed treaties.

These results have many implications for the study of international cooperation, institutions, and law. First, the mechanisms driving the tradeoff 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

[38] Amongst his many measures of the depth of cooperation, Haftel (2010) examines: monetary and fiscal cooperation, sectoral cooperation and harmonization, economic development, and provisions on the free movement of goods, services, capital, investment, and labor.

on the efforts of other states, particularly when states are under intense domestic pressure. Rigid enforcement systems can help to ameliorate this pressure and enhance compliance, but such systems also decrease stability by enhancing the likelihood of exit from the regime during tough times.

Second, my study challenges the reduction of legalization into a one-dimensional attribute, varying from "hard" to "soft" law.[39] "Hard law" is defined as a system of precise rules with strong delegation to third parties and a high degree of 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 depth-versus-rigidity 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 amongst treaty design attributes rather than restricting focus to a one-dimensional understanding of legalization.

Third, a basic tenet of political science is that political institutions affect policy outcomes. However, recent empirical studies of international institutions have found 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" (Rose 2004*a*, 98). He has also provided evidence that the GATT/WTO has no impact on the actual trade policies chosen by states or the volatility of international trade.[40] 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.[41] Nevertheless, they provoke distress for many: why do international institutions often appear to be ineffective in changing state behavior?

Past theoretical accounts have emphasized two factors that limit the effectiveness of international institutions: selection effects and exit options. Since a state chooses whether to be a member

[39] See Abbott and Snidal (2000) and Goldstein et al. (2000).

[40] See Rose (2004*b*) and Rose (2005).

[41] For challenges to Rose, see Goldstein, Rivers and Tomz (2007), Mansfield and Reinhardt (2008), Subramanian and Wei (2007), and Tomz, Goldstein and Rivers (2007). For challenges to Von Stein, see Simmons (2000) and Simmons and Hopkins (2005).

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, Rocke and Barsoom 1996, 380). Treaty regimes may provide screening effects that allow some states to cooperate while others are excluded.[42] 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 the level of cooperation and affect the distribution of costs and benefits within the membership.[43]

My theory offers an alternative and complimentary explanation: treaties that ostensibly require deep concessions 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 domestically. This means that the impact of institutions on international cooperation is itself a function of domestic politics. Changes in political and economic conditions at the domestic level 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.

[42] Von Stein (2005).
[43] See Johns (2007) and Voeten (2001).

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 noncompliance and not paying the fine is:

$$EU(D|t,a) \quad = \quad a\,u(t) - t - \int_0^A u(\tau(\alpha))\,dH(\alpha) + \int_{\alpha_S}^{\alpha_D} \sigma L\left(\tau\left(\alpha\right)\right) dH(\alpha) + \delta\chi_N$$

So the optimal defection tariff solves:

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

This constitutes noncompliance 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 noncompliance and paying the fine (settlement) is:

$$\begin{aligned} EU(S|t,a) \quad = \quad & a\,u(t) - t - \sigma L^*(t) - \int_0^A u(\tau(\alpha))dH(\alpha) + \int_{\alpha_S}^{\alpha_D} \sigma L(\tau(\alpha))dH(\alpha) \\ & + H\left(\alpha_D\right)\delta\beta\chi_C + \left[1 - H\left(\alpha_D\right)\right]\delta\chi_N \end{aligned}$$

So the optimal settlement tariff solves:

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

This constitutes noncompliance iff:

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

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 1. The home country's expected utility from actions C, S, and D given tariff levels from Lemma 1 are:

$$\begin{aligned}EU\left(C|t_B(a),a\right) &= a\,u\left(t_B(a)\right) - t_B(a) - \int_0^A u(\tau(\alpha))dH(\alpha) + \int_{\alpha_S}^{\alpha_D} \sigma L(\tau(\alpha))dH(\alpha) \\ &\quad + H\left(\alpha_D\right)\delta\beta\chi_C + \left[1 - H\left(\alpha_D\right)\right]\delta\chi_N \\ EU\left(S|t_S(a),a\right) &= a\,u\left(t_S(a)\right) - t_S(a) - \sigma L^*\left(t_S(a)\right) - \int_0^A u(\tau(\alpha))dH(\alpha) + \int_{\alpha_S}^{\alpha_D} \sigma L(\tau(\alpha))dH(\alpha) \\ &\quad + H\left(\alpha_D\right)\delta\beta\chi_C + \left[1 - H\left(\alpha_D\right)\right]\delta\chi_N \\ EU\left(D|t_D(a),a\right) &= a\,u\left(t_D(a)\right) - t_D(a) - \int_0^A u(\tau(\alpha))dH(\alpha) + \int_{\alpha_S}^{\alpha_D} \sigma L(\tau(\alpha))dH(\alpha) + \delta\chi_N\end{aligned}$$

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

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

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

$$\begin{aligned}\frac{\partial \hat{\Delta}}{\partial a} &= u\left(t_B\right) - u\left(t_S(a)\right) - (a-\sigma)\,u'\left(t_S(a)\right)\frac{\partial t_S(a)}{\partial a} + \frac{\partial t_S(a)}{\partial a} \\ &= u\left(t_B\right) - u\left(t_S(a)\right) < 0\end{aligned}$$

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

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

$$\begin{aligned}\text{So:}\qquad \frac{\partial \bar{\Delta}}{\partial a} &= (a-\sigma)\,u'\left(t_S(a)\right)\frac{\partial t_S(a)}{\partial a} - \frac{\partial t_S(a)}{\partial a} + u\left(t_S(a)\right) \\ &\quad + \frac{\partial t_D(a)}{\partial a} - a u'\left(t_D(a)\right)\frac{\partial t_D(a)}{\partial a} - u\left(t_D(a)\right) \\ &= u\left(t_S(a)\right) - u\left(t_D(a)\right) < 0\end{aligned}$$

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

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

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

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

As shown below in Proposition 3, a_D grows larger as β increases. Note that $a_S\left[u(t_B)-u(t_D(a_S))\right]$ is not a function of a_D. If the utility functions are such that $\frac{\partial}{\partial a}\left[t_D(a)-t_S(a)\right]\leq 0$, this is sufficient to show that $\overline{\Delta}(a_S)>0$ 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:

$$\chi_N = \frac{1}{1-\delta}\int_0^A\left[(a-1)u(t_D(a))-t_D(a)\right]dH(a)$$

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

$$\begin{aligned}\chi_C &= \int_0^A\left[au(t_E(a))-t_E(a)\right]dH(a)-\sigma\int_{a_S}^{a_D}L^*(t_E(a))\,dH(a)-\int_0^A u(\tau_E(\alpha))\,dH(\alpha)\\ &\quad +\sigma\int_{\alpha_S}^{\alpha_D}L^*(\tau_E(\alpha))\,dH(\alpha)+\delta H(a_D)^2\beta\chi_C+\delta\left[1-H(a_D)^2\right]\chi_N\\ &= \frac{\Psi}{1-\delta\beta H(a_D)^2}\\ \text{where}\quad \Psi &= \int_0^A\left[(a-1)u(t_E(a))-t_E(a)\right]dH(a)+\delta\left[1-H(a_D)^2\right]\chi_N\end{aligned}$$

Proof of Proposition 2. Recall that full compliance occurs if $a < a_S = \frac{1}{u'(t_B)} + \sigma$ and the probability of full compliance is $H(a_S)$.

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

□

Proof of Proposition 3. 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 $H(a_D)$ and $h(a_D)$ grow smaller. So for sufficiently large A:

$$\begin{aligned}
\lambda_{a_D} &= (a_D - \sigma) u'(t_S(a_D)) \frac{\partial t_S(a_D)}{\partial a_D} - \frac{\partial t_S(a_D)}{\partial a_D} - a_D u'(t_D(a_D)) \frac{\partial t_D(a_D)}{\partial a_D} + \frac{\partial t_D(a_D)}{\partial a_D} \\
&\quad + u(t_S(a_D)) - u(t_D(a_D)) + \delta H(a_D) \beta \frac{\partial \chi_C}{\partial a_D} + \delta h(a_D)(\beta \chi_C - \chi_N) \\
&= u(t_S(a_D)) - u(t_D(a_D)) + \delta H(a_D) \beta \frac{\partial \chi_C}{\partial a_D} + \delta h(a_D)(\beta \chi_C - \chi_N) < 0 \\
\lambda_{t_B} &= \sigma u'(t_B) + \delta H(a_D) \beta \frac{\partial \chi_C}{\partial t_B} > 0 \\
\lambda_\sigma &= (a_D - \sigma) u'(t_S(a_D)) \frac{\partial t_S(a_D)}{\partial \sigma} - \frac{\partial t_S(a_D)}{\partial \sigma} - L^*(t_S(a_D)) + \delta H(a_D) \beta \frac{\partial \chi_C}{\partial \sigma} \\
&= -L^*(t_S(a_D)) + \delta H(a_D) \beta \frac{\partial \chi_C}{\partial \sigma} < 0 \\
\lambda_\beta &= \frac{\delta H(a_D) \Psi}{\left[1 - \delta \beta H(a_D)^2\right]^2} > 0
\end{aligned}$$

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

□

Proof of Proposition 4. 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 $\chi_C(t_B, \sigma) = \chi^*$. The two first-order conditions for such an optimal pair (t_B, σ) are:

$$\begin{aligned}\frac{d\chi_C}{dt_B} &= \frac{\partial\chi_C}{\partial t_B}+\frac{\partial\chi_C}{\partial a_D}\frac{\partial a_D}{\partial t_B}=0\\ \frac{d\chi_C}{d\sigma} &= \frac{\partial\chi_C}{\partial\sigma}+\frac{\partial\chi_C}{\partial a_D}\frac{\partial a_D}{\partial\sigma}=0\end{aligned}$$

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 $\chi_C(t_B, \sigma) = \chi^*$:

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

where:

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

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

□

References

Abbott, Kenneth W. and Duncan Snidal. 2000. "Hard and Soft Law in International Governance." *International Organization* 54:421–456.

Alter, Karen J. and Laurence R. Helfer. 2010. "Nature or Nurture? Judicial Lawmaking in the European Court of Justice and the Andean Tribunal of Justice." *International Organization* 64:563–592.

Baccini, Leonardo, Andreas Dur, Manfred Elsig and Karolina Milewicz. 2011. "The Politics of Trade Agreement Design: Depth, Scope, and Flexibility." Paper presented at the 2011 Annual Conference of the International Political Economy Society.

Bagwell, Kyle and Robert W. Staiger. 1999. "An Economic Theory of the GATT." *American Economic Review* 89(1):215–248.

Baldwin, Richard. 1987. "Politically Realistic Objective Functions and Trade Policy." *Economics Letters* 24(3):287–90.

Busch, Marc L. and Eric Reinhardt. 2000. "Bargaining in the Shadow of the Law: Early Settlement in GATT/WTO Disputes." *Fordham International Law Journal* 24:158–72.

Busch, Marc L. and Krzysztof J. Pelc. 2010. "Tariff Bindings, Tariff Overhang, and Trade Remedies: Policy Flexibility at the WTO in Tough Political Times." Georgetown University.

Calvert, Randall L. 1995. Rational Actors, Equilibrium, and Social Institutions. In *Explaining Social Institutions*, ed. Jack Knight and Itai Sened. University of Michigan Press pp. 57–93.

Dai, Xinyuan. 2005. "Why Comply? The Domestic Constituency Mechanism." *International Organization* 59:363–98.

Downs, George W. and David M. Rocke. 1995. *Optimal Imperfection? Domestic Uncertainty and Institutions in International Relations*. Princeton University Press.

Downs, George W., David M. Rocke and Peter N. Barsoom. 1996. "Is the Good News about Compliance Good News about Cooperation?" *International Organization* 50:379–406.

Fearon, James D. 1998. "Bargaining, Enforcement and International Cooperation." *International Organization* 52:269–306.

Fudenberg, Drew and Jean Tirole. 2000. *Game Theory.* MIT Press.

Gilligan, Michael. 1997. *Empowering Exporters: Delegation, Reciprocity and Collective Action in Twentieth Century American Trade Policy.* University of Michigan Press.

Gilligan, Michael, Leslie Johns and B. Peter Rosendorff. 2010. "Strengthening International Courts and the Early Settlement of Disputes." *Journal of Conflict Resolution* 54:5–38.

Goldstein, Judith L., Douglas Rivers and Michael Tomz. 2007. "Institutions in International Relations: Understanding the Effects of the GATT and the WTO on World Trade." *International Organization* 61:37–67.

Goldstein, Judith and Lisa L. Martin. 2000. "Legalization, Trade Liberalization, and Domestic Politics: A Cautionary Note." *International Organization* 54:603–632.

Goldstein, Judith, Miles Kahler, Robert O. Keohane and Anne-Marie Slaughter. 2000. "Introduction: Legalization and World Politics." *International Organization* 54:385–99.

Grossman, Gene M. and Elhanan Helpman. 1994. "Protection for Sale." *American Economic Review* 84:833–850.

Haftel, Yoram Z. 2010. "Commerce and Institutions: Trade, Scope, and the Design of Regional Economic Organizations." Working Paper. Presented at the Workshop on the Politics of Trade Agreements: Theory, Measurement, and Empirical Applications. April 30-May 1, 2010, Princeton University.

Hicks, Raymond and Soo Yeon Kim. 2010. "Credible Commitment through PTAs and their Effects on Trade: A Study of Asia's Reciprocal Trade Agreements." Working Paper.

Jackson, John H. 2004. "International Law Status of WTO Dispute Settlement Reports: Obligations to Comply or Options to 'Buy Out'?" *American Journal of International Law* 98:109–125.

Johns, Leslie. 2007. "A Servant of Two Masters: Communication and the Selection of International Bureaucrats." *International Organization* 61:245–275.

Johns, Leslie. 2011. "Strengthening International Courts." UCLA Book Manuscript.

Johns, Leslie. 2012. "Courts as Coordinators: Endogenous Enforcement and Jurisdiction in International Adjudication." *Journal of Conflict Resolution* pp. XX–XX.

Johns, Leslie and B. Peter Rosendorff. 2009. Dispute Settlement, Compliance and Domestic Politics. In *Trade Disputes and the Dispute Settlement Understanding of the WTO: An Interdisciplinary Assessment.*, ed. James C. Hartigan. Emerald Group Publishing pp. 139–163.

Kucik, Jeffrey. 2011. "The Domestic Politics of International Institutional Design: Industry Preferences over Preferential Trade Agreements." Working Paper.

Kucik, Jeffrey and Eric Reinhardt. 2008. "Does Flexibility Promote Cooperation? An Application to the Global Trade Regime." *International Organization* 62:477–505.

Maggi, Giovanni and Robert W. Staiger. 2010. "Breach, Remedies and Settlement in Trade Agreements." Working Paper.

Mansfield, Edward D. and Eric Reinhardt. 2008. "International Institutions and the Volatility of International Trade." *International Organization* 62:621–652.

Mansfield, Edward D., Helen V. Milner and B. Peter Rosendorff. 2000. "Free to Trade: Democracies, Autocracies and International Trade." *American Political Science Review* 94:305–322.

Milgrom, Paul R., Douglass C. North and Barry R. Weingast. 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, Helen V., Edward D. Mansfield and B. Peter Rosendorff. 2004. International Trade and Domestic Politics: The Domestic Sources of International Trade Agreements and Institutions. In *The Impact of International Law on International Cooperation: Theoretical Perspectives*, ed. Eyal Benvenisti and Moshe Hirsch. Cambridge: Cambridge University Press.

Rose, Andrew K. 2004*a*. "Do We Really Know That the WTO Increases Trade?" *American Economic Review* 94:98–114.

Rose, Andrew K. 2004*b*. "Do WTO Members Have More Liberal Trade Policy?" *Journal of International Economics* 63:209–235.

Rose, Andrew K. 2005. "Does the WTO Make Trade More Stable?" *Open Economies Review* 16:7–22.

Rosendorff, B. Peter. 2005. "Stability and Rigidity: Politics and the Design of the WTO's Dispute Resolution Procedure." *American Political Science Review* 99:389–400.

Rosendorff, B. Peter and Helen V. Milner. 2001. "The Optimal Design of International Institutions: Uncertainty and Escape." *International Organization* 55:829–857.

Rubinstein, Ariel. 1991. "Comments on the Interpretation of Game Theory." *Econometrica* 59:909–924.

Schwartz, Warren F. and Alan Sykes. 2002. "The Economic Structure of Renegotiation and Dispute Settlement in the World Trade Organization." *Journal of Legal Studies* 31:S179–S204.

Simmons, Beth. 1998*a*. "Compliance with International Agreements." *Annual Review of Political Science* 1:75–93.

Simmons, Beth. 2000. "International Law and State Behavior: Commitment and Compliance in International Monetary Affairs." *American Political Science Review* 94:819–835.

Simmons, Beth A. 1998*b*. "Compliance with International Agreements." *Annual Review of Political Science* 1:75–93.

Simmons, Beth A. 2009. *Mobilizing for Human Rights: International Law in Domestic Politics*. Cambridge University Press.

Simmons, Beth A. and Daniel J. Hopkins. 2005. "The Constraining Power of International Treaties: Theory and Methods." *American Political Science Review* 99:623–631.

Smith, James McCall. 2000. "The Politics of Dispute Settlement Design: Explaining Legalism in Regional Trade Pacts." *International Organization* 54:137–180.

Subramanian, Arvind and Shang-Jin Wei. 2007. "The WTO Promotes Trade, Strongly but Unevenly." *Journal of International Economics* 72:151–175.

Tomz, Michael, Judith L. Goldstein and Douglas Rivers. 2007. "Do We Really Know that the WTO Increases Trade? Comment." *American Economic Review* 97:2005–2018.

Voeten, Eric. 2001. "Outside Options and the Logic of Security Council Action." *American Political Science Review* 95:845–58.

Von Stein, Jana. 2005. "Do Treaties Constrain or Screen? Selection Bias and Treaty Compliance." *American Political Science Review* 99:611–622.

Figure 1: Equilibrium Behavior

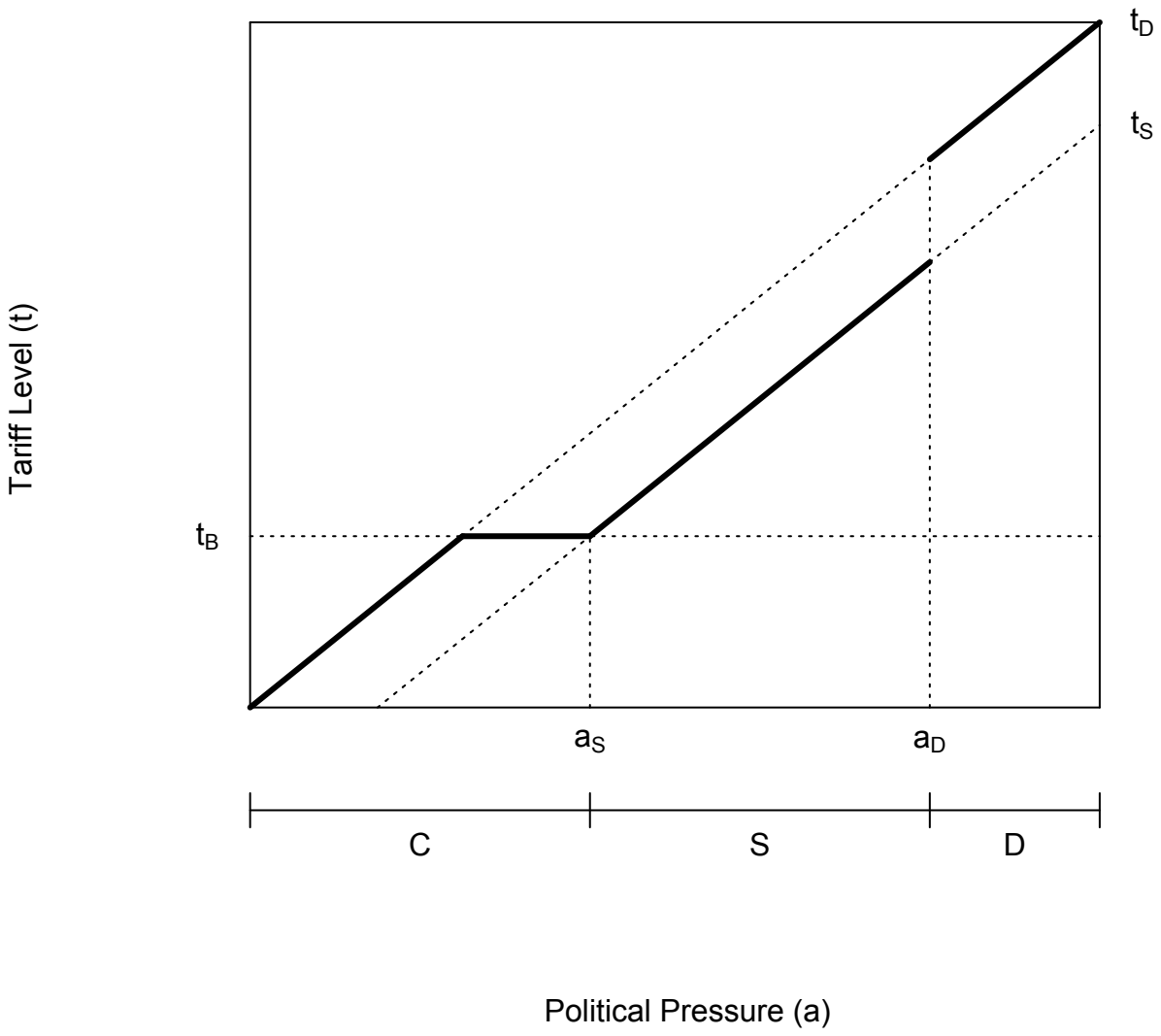

Figure 2: The Impact of Deepening Tariff Concessions

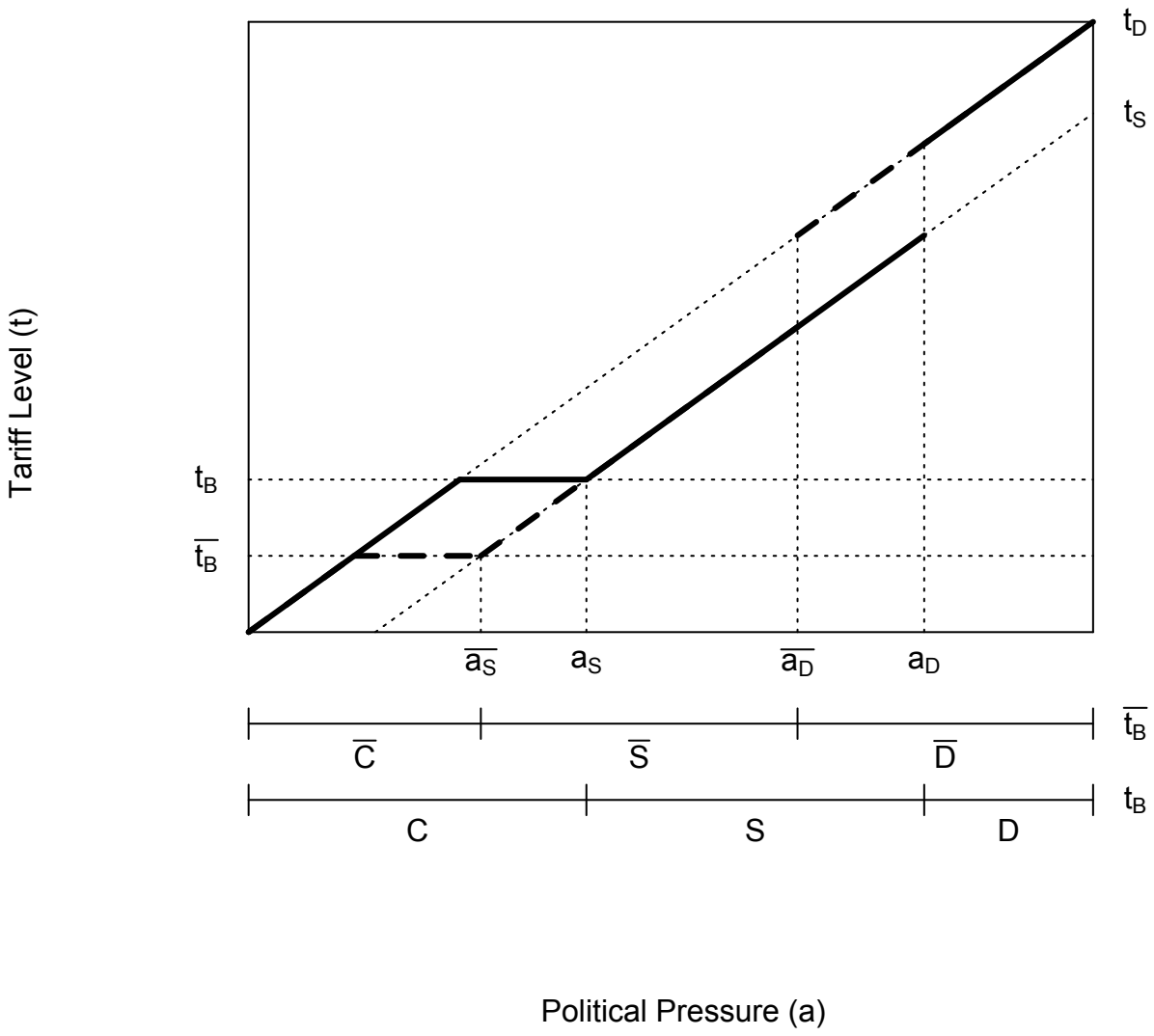

Figure 3: The Impact of Increasing Rigidity