

# Self-enforcing Trade Agreements, Dispute Settlement and Separation of Powers

Kristy Buzard

*110 Eggers Hall, Economics Department, Syracuse University, Syracuse, NY 13244. 315-443-4079.*

---

## Abstract

In an environment where international trade agreements must be enforced via promises of future cooperation, the presence of an import-competing lobby has important implications for optimal punishments, and therefore dispute resolution procedures. When lobbies work to disrupt trade agreements, a Nash reversion punishment scheme must balance two, conflicting objectives. Longer punishments help to enforce cooperation by increasing the government's costs of defecting, but because the lobby prefers the punishment outcome, this also incentivizes lobbying effort and with it political pressure to break the agreement. Thus the model generates new predictions for the design of mechanisms for resolving trade disputes: there is an optimal length for dispute resolutions procedures, and it depends directly on the political influence of the lobbies. Trade agreement tariffs are shown to be increasing in the political influence of the lobbies, as well as their patience levels.

*Keywords:*

trade agreements, lobbying, WTO, dispute settlement, repeated games, enforcement

---

## 1. Introduction

In the absence of strong external enforcement mechanisms for international trade agreements, we generally assume that cooperation is enforced by promises of future cooperation, or, equivalently, promises of future punishment for exploitative behavior.

---

*Email address:* kbuzard@syr.edu (Kristy Buzard)

*URL:* <http://faculty.maxwell.syr.edu/kbuzard> (Kristy Buzard)

When repeated-game incentives are used to enforce cooperation and prevent players from defecting in a prisoner's dilemma-style stage game, the strongest punishment available is usually assumed to be the grim trigger strategy of defecting forever upon encountering a defection by one's partner.

I show that when lobbies are relevant players in the repeated game, the optimal length of Nash-reversion punishments is finite and can be derived directly from the players' incentive constraints.<sup>1</sup> In the context of international trade agreements, these punishments can be interpreted as arising from the design and implementation of rules like those established in the World Trade Organization's Dispute Settlement Understanding.

Not only does adding lobbies suggest an optimal length for punishments that feature periods of Nash-reversion-style non-cooperation; it also turns out that this optimal punishment length itself depends on how readily special interests are able to influence the political process. That is, the optimal length of the dispute settlement procedure is a function of the strength of the lobbies, reinforcing the idea that including lobbying in such analyses can be critically important for institutional design questions. We shall also see that, for a fixed dispute settlement procedure, as lobbies become more influential or more patient, the equilibrium trade agreement tariff that must be provided in order to overcome the ratification hurdle increases.

The structure of the model is similar to that of ? with two main changes: the political-economy weights are endogenously determined, and in place of a unitary government that has different preferences before and after signing a trade agreement, this model has two branches of government with differing preferences who share policy-making power as in ?, ? and ?.

This paper incorporates such a separation-of-powers policy-making process with endogenous lobbying along the lines of ?? into a repeated-game setting. Here, welfare-

---

<sup>1</sup>There are results in the literature—e.g. ?—that involve finite-length optimal punishments. To the best of my knowledge, all of these are for environments in which the players spend some time in the punishment phase, usually due to imperfect monitoring and/or uncertainty. Thus the shortening of the punishment serves to increase welfare by minimizing time spent in punishment periods. The results of this paper are of a different nature, as the players remain in the cooperative state in all periods. Here, the gain comes from loosening a player's incentive constraint so that cooperative state welfare is higher.

maximizing executives use their control over trade-agreement tariffs as a kind of political commitment device:<sup>2</sup> by setting tariffs to optimally reduce lobbying incentives, they reduce the political pressure on the legislatures. This changes the legislatures' incentives, so that they do not break the agreement as they would have if they had faced more intense political pressure.<sup>3</sup>

Given that all actors have perfect information about the effect of lobbying effort on the outcome of the political process, the executives maximize social welfare by choosing the lowest tariffs that make it unattractive for the lobbies to provoke the legislature to initiate a trade dispute.<sup>4</sup> So even when there are no disputes in equilibrium, the out-of-equilibrium threat that a lobby might provoke a trade war is crucial in determining the equilibrium trade agreement structure.

Thus the problem with the lobby has an extra constraint relative to the standard problem. The constraint on the key repeated-game player, which for simplicity is described herein as the legislature, is loosened by increasing the punishment length because defections become relatively more unattractive. However, the new constraint due to the presence of lobbying becomes tighter as the punishment becomes more severe because the lobby *prefers* punishment periods. Because the tariffs during punishment, and thus the lobby's profits, are higher compared to those they receive during a cooperative period, the lobby has increased incentive to exert effort as the punishment lengthens.

The optimal punishment length must balance these two competing forces. Where

---

<sup>2</sup>This is a different kind of domestic commitment role for trade agreements than that identified by ?, who show that trade agreements can be useful for helping governments commit vis-à-vis private firms in their investment decisions.

<sup>3</sup>This is not to say that the legislature itself is made better off by the reduction in political pressure, although ? demonstrates that this is possible. It only means that the executive can use the commitment power of the trade agreement to improve its welfare, which is assumed to differ from that of the legislature. The commonly-made assumption that the executive is less protectionist than the legislature is a special case of the finding that susceptibility to special interests generally declines with the size of one's constituency. One simple illustration from the realm of trade policy is the following: a legislator whose district has a large concentration of a particular industry does not take into account the impact of tariffs on the welfare of consumers in other districts, while the executive, whose constituency encompasses the whole country, will internalize these diffuse consumption effects. For a detailed argument, see ?.

<sup>4</sup>With no uncertainty of any kind, there will be no trade disputes in equilibrium. Political uncertainty can be easily added to the model, in which case lobbying effort is typically non-zero and there is a positive probability of dispute in equilibrium.

the balance falls depends in large part on how influential the lobby is in the legislative process. If the lobby has very little power, the optimal punishment converges to that of the model without a lobby: longer punishments are better because the key constraint is the legislature's. As the lobby becomes stronger, the optimal punishment becomes shorter because the lobby's incentive becomes more important.

Quite intuitively, it is also shown that, for a given punishment length, increases in the lobby's strength lead to lower required payments to provoke trade dispute and therefore higher equilibrium trade agreement tariffs to avoid those disputes. Increases in the lobby's patience have the same qualitative effects, while increases in the patience of the legislature work in the opposite direction: the lobby must pay more to induce the legislature to endure the punishment and the executive can accordingly reduce trade agreement tariffs without fear that they will be broken.

Repeated non-cooperative game models of trade agreements have been considered by ??, ?, ?, ?????, ?, ?, ?, ?, ?, ?, and ?.

In particular, ?, ?, ?, ? and ? consider the impact of different assumptions about reactions and timing of punishments for deviations from agreements. Here, I study a very simple structure in which the trading partners remain in a symmetric trade war for a predetermined number of periods.

The length of this punishment phase could be interpreted as being determined by the specific dispute resolution procedures of an institution such as the World Trade Organization.<sup>5</sup> The model would require some modifications in order to match a multi-lateral agreement with many goods, for instance specifying that trade goes on as usual in all those industries except the one in which the applied tariff is raised above the tariff cap and the industry the trading partner chooses to use for retaliation. But the basic intuition goes through: the incentives of lobbies should be taken into account when designing dispute resolution procedures because the length of time a lobby can expect to enjoy a higher trade-war tariff is directly related to whether the lobby finds it

---

<sup>5</sup>The model can also be applied to Preferential Trade Agreements with some additional modifications due to the restrictions imposed by GATT Article XXIV. Since the interpretation of the restriction that PTAs cover 'substantially all the trade' has never been settled in law, there remains significant scope to grant non-zero tariffs to industries who exert sufficient lobbying effort.

worthwhile to exert effort in provoking a dispute in the first place.

In line with this fundamental idea, I show that a punishment scheme that involves the defecting party applying a zero tariff during the punishment can support lower trade agreement tariffs than does reverting to the stage-game Nash equilibrium. This occurs because these low tariffs significantly weaken the lobby's incentive to exert effort to break the trade agreement. I am not aware of such punishments being applied in actual trade agreements, and this may be because other considerations rule out this type of punishment. But it is worth considering whether some such alternative punishment structure that takes into account lobbying incentives may be implementable and thus capable of supporting greater levels of cooperation.

The model under consideration here can only speak directly to motives for pure rent-seeking and not to responses to unpredictable changes in the economic and political environment since such uncertainty is assumed away. This means that measures designed to provide escape are not beneficial in this environment (cfr. ?, ?). With no uncertainty, disputes should not be observed on the equilibrium path. In reality, of course, there is considerable such uncertainty, but it's not clear that this is the sole source of the trade disputes that arise.

For instance, the immediate retaliation that ensures self-enforcement in this model is rarely possible under current trading rules and this may well increase the number of disputes observed in equilibrium. One possibility for implementing more immediate retaliation is the idea proposed in the literature that trading partners exact 'vigilante justice' through various means such as imposing unrelated anti-dumping duties.<sup>6</sup> However, this would not necessarily reduce the number of disputes if the original defector objects to the new anti-dumping measure. In order for the 'vigilante justice' option to work as a punishment in the context of this model, the original defector would have to tacitly acknowledge it as punishment and play along.

I begin in the next section by describing in detail the stage game, which is closely related to the model in ?. Both papers employ the separation-of-powers government structure with endogenous lobbying. While the current paper focuses on the implica-

---

<sup>6</sup>See the discussions in ? and ? for evidence on informal versus formal retaliation.

tions of self-enforcement constraints for the optimal design of trade agreements and dispute-settlement institutions, ? abstracts from enforcement issues and demonstrates that taking into account the separation-of-powers structure can resolve the empirical puzzle surrounding the ? ‘Protection for Sale’ model, highlights the importance of the threat of ratification failures on the formation of trade agreements and develops new results about the role of political uncertainty in the policy-making process.

Section 3 then sets out the dispute settlement institution and the set-up of the repeated game. I describe the structure and properties of trade agreements in this environment in Section 4 and explore the forces shaping dispute resolution procedures in Section 5. Section 6 supplies the details for the repeated game enforcement and explores an alternative punishment scheme. Section 7 concludes.

## 2. Stage Game

I employ a two-good partial equilibrium model with two countries: home (no asterisk) and foreign (asterisk). The countries trade two goods,  $X$  and  $Y$ , where  $P_i$  denotes the home price of good  $i \in \{X, Y\}$  and  $P_i^*$  denotes the foreign price of good  $i$ . In each country, the demand functions are taken to be identical for both traded goods, respectively  $D(P_i)$  in home and  $D(P_i^*)$  in foreign and are assumed strictly decreasing and twice continuously differentiable.

The supply functions for good  $X$  are  $Q_X(P_X)$  and  $Q_X^*(P_X^*)$  and are assumed strictly increasing and twice continuously differentiable for all prices that elicit positive supply. I also assume  $Q_X^*(P_X) > Q_X(P_X)$  for any such  $P_X$  so that the home country is a net importer of good  $X$ . The production structure for good  $Y$  is taken to be symmetric, with both demand and supply such that the economy is separable in goods  $X$  and  $Y$ . As is standard, it is assumed that the production of each good requires the possession of a sector-specific factor that is available in inelastic supply and is non-tradable so that the income of owners of the specific factors is tied to the price of the good in whose production their factor is used.

For simplicity, I assume each government’s only trade policy instrument is a specific tariff on its import-competing good: the home country levies a tariff  $\tau$  on good  $X$  while the foreign country applies a tariff  $\tau^*$  to good  $Y$ . Local prices are then

$P_X = P_X^W + \tau$ ,  $P_X^* = P_X^W$ ,  $P_Y = P_Y^W$  and  $P_Y^* = P_Y^W + \tau^*$  where a  $W$  superscript indicates world prices and equilibrium prices are determined by the market clearing conditions

$$M_X(P_X) = D(P_X) - Q_X(P_X) = Q_X^*(P_X^*) - D(P_X^*) = E_X^*(P_X^*)$$

$$E_Y(P_Y) = Q_Y(P_Y) - D(P_Y) = D(P_Y^*) - Q_Y^*(P_Y^*) = M_Y^*(P_Y^*)$$

where  $M_X$  are home-county imports and  $E_X^*$  are foreign exports of good  $X$  and  $E_Y$  are home-county exports and  $M_Y^*$  are foreign imports of good  $Y$ .

It follows that  $P_X^W$  and  $P_Y^W$  are decreasing in  $\tau$  and  $\tau^*$  respectively, while  $P_X$  and  $P_Y^*$  are increasing in the respective domestic tariff. This gives rise to a standard terms-of-trade externality. As profits and producer surplus (identical in this model) in a sector are increasing in the price of its good, profits in the import-competing sector are also increasing in the domestic tariff. This economic fact, combined with the assumptions on specific factor ownership, is what motivates political activity.

I next describe the politically-relevant actors. In order to focus attention on protectionist political forces, I assume that only the import-competing industry in each country is politically-organized and able to lobby and that it is represented by a single lobbying organization.<sup>7</sup> Each country's government is composed of two branches: an executive who can conclude trade agreements and a legislature that has final say on trade policy. Thus the political process is modeled as involving three players in each country: the lobby, the executive, and the legislature.

The stage-game timing is as follows. At time zero, the executives set trade policy cooperatively in an international agreement. The agreement takes the form of a cap on tariffs so that if the applied tariff in any period is higher than the agreed-upon level, there is a breach of the agreement.

In each subsequent period, each lobby decides whether to exert the effort to convince the legislature in its respective country to break the trade agreement or to exert the lower effort level that achieves the trade agreement tariff. Depending on lobbying

---

<sup>7</sup>Adding a pro-trade lobby for the exporting industry would modify the magnitude of the effects and make free trade attainable for a range of parameter values, but it would not modify the essential dynamic.

effort, the legislatures then decide whether to abide by the agreement or to provoke a trade war.

In the event that at least one of the legislatures chooses a tariff greater than the trade agreement tariff, the trade agreement tariffs are suspended and the relationship moves into a punishment phase. The following period has a different structure: there is no trade agreement to break, so there is simply lobbying to influence the level at which the punishment phase tariff is set and a decision on these tariffs by the legislatures. Once all political decisions are taken, producers and consumers make their decisions.

We are in an environment of complete information, so the appropriate solution concept is subgame perfect Nash equilibrium. Whenever this is a possibility of multiple equilibria, I will focus on the one that maximizes the welfare of the executives. As this game is solved by backward induction, it is intuitive to start by describing the incentives of the legislatures, whose decisions I model as being taken by a median legislator. As the economy is fully separable and the economic and political structures are symmetric, I focus here on the home country and the  $X$ -sector. The details are analogous for  $Y$  and foreign.

The per-period welfare function of the home legislature is

$$W_{ML} = CS_X(\tau) + \gamma(e) \cdot \pi_X(\tau) + CS_Y(\tau^*) + \pi_Y(\tau^*) + TR(\tau) \quad (1)$$

where  $CS$  is consumer surplus,  $\pi$  are profits (identical to producer surplus in this model),  $\gamma(e)$  is the weight placed on profits in the import-competing industry,  $e$  is lobbying effort, and  $TR$  is tariff revenue. Here, the weight the median legislator places on the profits of the import-competing industry,  $\gamma(e)$  is affected by the level of lobbying effort.

Notice that, aside from the endogeneity of the weight the legislature places on the lobbying industry's profits, this is precisely the *deus ex machina* government objective function popularized by ? that is commonly employed in the literature on the political economy of trade agreements. Since trade policies are often determined within the context of trade agreements, it is useful to have a framework to bring together the endogenous political pressure of 'Protection-for-Sale'-style modeling with the trade agreements approach; the formulation in Equation 1 is intended to be a bridge between



the two.

In the literature that studies the design of trade agreements and institutions, political pressure is taken to exogenously impact the value politicians place on producer surplus. Here, that level of political pressure is taken to be determined by lobbying effort, which can be interpreted broadly as any action that serves to increase the weight that the median legislator places on producer surplus when taking decisions. Modeling the objective function so closely on the standard in the trade agreements literature allows for direct comparisons to the large extant body of work that studies exogenous shocks only, revealing cleanly the effects of the addition of endogenous lobbying.

**Assumption 1.**  $\gamma(e)$  is continuously differentiable, strictly increasing and concave in  $e$ .

Assumption 1 formalizes the intuition that the legislature favors the import-competing industry more the higher is its lobbying effort, but that there are diminishing returns to lobbying activity.<sup>8</sup> The assumption of diminishing returns to lobbying effort has been present in the literature going back at least to ?. ? point out the linearity in contributions assumed in the Protection for Sale model prevents complete analysis of distributional questions and restricts the returns to lobbying activity to be constant.

The functional form in Expression 1 with Assumption 1 can be interpreted as a special case of the general welfare function proposed in ? in which the median legislature's welfare exhibits decreasing returns to lobbying effort.<sup>9</sup> The interpretation is that the identity of the median legislator changes ever more slowly as lobbying effort increases because it becomes more difficult for the lobby to win additional votes given that the most friendly legislators are targeted first.

Lobbying affects only the weight the legislature places on the profits of the import-competing industry. These profits are higher in a trade war than under a trade agreement, so given Assumption 1, the legislature becomes more favorably inclined toward the high trade-war tariff and associated profits as lobbying increases and therefore more

---

<sup>8</sup>The diminishing returns here take the form of declining increments to the lobby's influence as effort increases; in ?, the returns to lobbying decline with higher levels of protection.

<sup>9</sup>Note that while the model of ? nests both the model presented in this paper and that of ?, neither of the latter two are generalizations of the other. Although complex, an isomorphism can be made between the latter two in a special case as discussed in ?.

likely to break the trade agreement.

Given the legislature's preferences, the home lobby chooses its lobbying effort to maximize its net profits:

$$U_L = \pi(\tau) - e \quad (2)$$

where  $\pi(\cdot)$  is the current-period profit and  $\tau$  is the home country's tariff on the import good. Values of  $\tau$  that are of particular interest are  $\tau^a$ ,  $\tau^{tw}$ ,  $\tau^p$  and  $\tau^b$ , respectively the trade agreement, trade war, punishment and break tariffs. The corresponding effort levels are notated as  $e_a$  to influence the applied tariff under the trade agreement,  $e_{tw}$  in the trade war,  $e_p$  to influence the punishment tariff and  $e_b$  to influence the break decision. I use the convention throughout of representing a vector of tariffs for both countries  $(\tau, \tau^*)$  as a single bold  $\tau$ .

I assume the lobby's contribution is not observable to the foreign legislature. The implication is that the lobby can directly influence only the home legislature, and so the influence of one country's lobby on the other country's legislature occurs only through the tariffs selected.<sup>10</sup>

In the first stage, the executives choose the trade agreement tariffs  $\tau^a = (\tau^a, \tau^{*a})$  via a negotiating process that I assume to be efficient. In this symmetric environment, this process maximizes the joint payoffs of the trade agreement:<sup>11</sup>

$$W_E(\tau^a) = W_E(\tau^a) + W_E^*(\tau^a) \quad (3)$$

I model the executives' choice via the Nash bargaining solution where the disagreement point is the executives' welfare resulting from the Nash equilibrium in the non-cooperative game (i.e. in the absence of a trade agreement) between the legislatures.

The executives are assumed, for simplicity, to be social-welfare maximizers who can make transfers between them.<sup>12</sup> Therefore the home executive's welfare is speci-

---

<sup>10</sup>cfr. ?, page 685-686.

<sup>11</sup>If political uncertainty is present, the joint payoffs must take into account the possibility that the trade agreement will be broken. In the case of certainty, agreement will always be maintained on the equilibrium path and so this specification is sufficient.

<sup>12</sup>It is trivial to relax the assumption of social-welfare maximizing executives; in the present symmetric environment with no disputes, the same is true of the assumption about transfers.

fied as follows:

$$W_E = CS_X(\tau) + \pi_X(\tau) + CS_Y(\tau^*) + \pi_Y(\tau^*) + TR(\tau) \quad (4)$$

Note that this is identical to the welfare function for the legislature aside from the weight on the profits of the import industry, which is not a function of lobbying effort and here is assumed to be 1 for simplicity.

This stylized modeling of objective functions can accommodate real-world institutions such as those in the United States where the Congress has some consultative role in trade agreement negotiations and the executive branch has the ability to alter applied tariffs under important administrative procedures such as anti-dumping and safeguard measures.<sup>13</sup> One need only alter the interpretation of Equations 1 and 4 as the objectives of the government more broadly at the trade agreement and applied-tariff-setting stages respectively.

The idea is that lobbying has less of an impact during trade agreement negotiations—embodied in the executive’s objective function—than it does during day-to-day trade policy making, which is embodied in the legislature’s objective function. This set-up represents the difference between the impact of lobbying during the two stages in a simple, albeit extreme, way that permits a focus on the out-of-equilibrium threat of trade disruption created by the lobbies.<sup>14</sup>

Even this stark assumption does *not* require that there is no lobbying at the trade agreement stage. It only implies that the government’s preferences at this stage are not directly altered in a significant way by lobbying over trade. It is very reasonable to imagine that the executive takes into account information it gathers from lobbies in the ex-ante stage about how hard they will work to disrupt the trade agreement for a given level of tariffs.

---

<sup>13</sup>It is, however, debatable whether many of these procedures fall under the scope of the issues considered in this paper since they are often WTO-legal and therefore do not serve to violate the trade agreement. In any case, the conditions under which these procedures would be necessary in a trade agreement—where subsidies are a policy choice (countervailing duties), there is uncertainty about the trading environment (escape clause) or markets are not perfectly competitive (anti-dumping)—are not present in the environment under consideration here.

<sup>14</sup>The model is amenable to adding lobbying at the trade-agreement formation stage. This adds an interesting question of how lobbies make a resource allocation decision between the two stages. This is left for future work.

For trade policy, where there are concentrated benefits but harm is diffuse, there are good reasons for the legislature to be more protectionist than the president, as has been the case in the post-war United States. Because the President has the largest constituency possible, delegating authority to the executive branch may simply be a mechanism for “concentrating” the benefits since consumers seem unable to overcome the free-riding problem. In fact, a strong argument can be made that power over trade policy has been delegated to the executive branch precisely *because* it is less susceptible to the influence of special interests (Destler 2005).

Therefore, in line with both the theoretical and empirical literature, I will assume that even for the least favorable outcome of the lobbying process, the legislature will be at least weakly more protectionist than the executive.

**Assumption 2.**  $\gamma(e) \geq 1 \ \forall e$ .

Assumption 2 ensures that the trade agreement tariff is less than the tariff that results from unconstrained interaction between the lobby and legislature, which I denote  $\tau^{tw}$  and explain below. More generally, it guarantees that the legislature’s incentives are more closely aligned with the lobby’s than are those of the executive. This is not essential but simplifies the analysis and matches well the empirical findings that politicians with larger constituencies are less sensitive to special interests (See ? and footnote 3 above).

Although the political process here matches most closely that of the United States in the post-war era, I believe the model or one of its extensions is applicable for a broad range of countries for which authority over the formation and maintenance of trade policy is diffuse and subject to political pressure either at home or in a trading partner.<sup>15</sup>

### 3. Repeated Game

This trade policy environment has many features of a standard prisoner’s dilemma. Most importantly, the legislatures face unilateral incentives to violate the terms of the

---

<sup>15</sup>In particular, the binary decision by the legislature about whether to abide by or break the trade agreement is modeled on the “Fast Track Authority” that the U.S. Congress granted to the Executive branch almost continuously from 1974-1994 and then again as “Trade Promotion Authority” from 2002-2007.

agreement under pressure from the lobbies. When the legislatures and lobbies set tariffs at the higher stage-game Nash equilibrium level, payoffs for the social-welfare conscious executives are reduced. In order to maintain the trade agreement without external enforcement, we turn to repeated game incentives.

Precise conditions for supporting the trade agreement in equilibrium depend on the full set of punishments that are specified, and the details for several possibilities are given in Section 6. For the results in Sections 4 and 5, deviations from the trade agreement are punished by reverting to the stage game Nash equilibrium for a specified number of periods before returning to cooperation. This represents a trade war that is limited in duration and therefore more realistic than infinite Nash reversion. I show in Section 5 that infinite Nash reversion punishments are dominated by shorter lengths of Nash reversion when lobbies are non-trivial political actors.

For Nash reversion punishments of length  $T$ , the timing of actions is the following. If both countries apply tariffs less than or equal to the trade agreement level in period  $t$ , they continue to cooperate under the terms of the trade agreement in period  $t + 1$ . If one or both countries defect from the agreement in period  $t$  by applying a tariff greater than the trade agreement level, a dispute arises. In this case, they enter into an applied tariff stage for the  $T$  periods beginning in period  $t + 1$ . In each of these periods, the punishment specifies that the lobbies and legislatures in both countries play their stage-game Nash equilibrium strategies. In period  $t + T + 1$ , the trade agreement is reinstated.

In order to determine repeated-game incentives, we must fully specify the punishments for both on- and off-equilibrium path actions. We can think of the punishment scheme as being designed either by the executives or by a supranational body like the WTO. After the punishment scheme is designed, the executives choose the trade agreement tariffs to jointly maximize social welfare. They have no other opportunity to affect the outcome of the trading relationship. Thus the executives maximize joint welfare subject to the incentive constraints of the other players.

I assume that the punishment scheme and trade agreement are designed so that the legislatures and lobbies behave in such a way that the trade agreement remains in force in equilibrium. In Section 5, the question of how to optimally design the punishment

scheme within this class of  $T$ -period Nash reversion punishments is addressed.

Again, because of symmetry and separability, it suffices to restrict attention to the home country. An overview of the punishment scheme is given here; for details and a precise description of equilibrium strategies, see Section 6.

On the equilibrium path, the lobby chooses effort level  $e_a$ , which is less than a cut-off effort level  $\bar{e}$  that would cause the legislature to break the trade agreement.  $e_a$  is the effort level at which the legislature chooses  $\tau^a$  as its optimal unilateral tariff. The determination of  $\bar{e}$  is described in the next section. The legislature never breaks the trade agreement, with an applied tariff equal to the trade agreement tariff  $\tau^a$ . These choices are made as long as any violation of the agreement was at least  $T$  periods in the past.

Whenever there is a violation of the agreement—that is, a tariff higher than the trade agreement level—the lobby and the legislature play their static Nash strategies for  $T$  periods.

For the legislature, the Nash outcome is the tariff that unilaterally maximizes Equation 1 given  $\tau^*$  and the lobby's choice  $e_p$ . The separability of the economy implies that there are no cross-country interactions in the decision problems, so the home and foreign best response tariffs are independent and the home country's tariff in a punishment period maximizes weighted home-country welfare in the  $X$ -sector only. The foreign legislature's decision problem is analogous, and unilateral optimization leads to what I refer to as the trade war tariffs  $\tau^{tw}$  as the solution to the following first order condition:<sup>16</sup>

$$\frac{\partial CS_X(\tau)}{\partial \tau^p} + \gamma(e_p) \cdot \frac{\partial \pi_X(\tau)}{\partial \tau^p} + \frac{\partial TR(\tau)}{\partial \tau^p} = 0$$

The lobby chooses its effort  $e_p$  given the above best response tariff-setting behavior denoted by  $\tau^R$  by maximizing its profits net of effort:  $\pi(\tau^R(\gamma(e_p))) - e_p$ . This implies a first order condition of

$$\frac{d\pi(\tau^R(\gamma(e_p)))}{de_p} = 1 \quad (5)$$

---

<sup>16</sup>That the second order condition is satisfied is not guaranteed. See the appendix of ? for a discussion as well as a sufficient condition when prices are linear in tariffs. At issue is the need to bound the impact of the convexity of the profit term relative to the concavity of the consumer surplus term for any given value of  $\gamma$ .

That is, at this stage, the lobby chooses the level of effort that equates its expected marginal increase in profits with its marginal payment. I label this effort level  $e_{tw}$  because  $\tau^R(\gamma(e_{tw})) = \tau^{tw}$ , the trade war tariff.<sup>17</sup>

Unlike with infinite Nash reversion punishments, these  $T$ -period Nash reversion punishments are not necessarily subgame perfect since the players may want to influence the future path of play during punishment periods. Subgame perfection can be ensured by punishing any deviations *from the punishment* by restarting either the punishment for the legislature or the trade agreement for the lobby. This works because the legislature prefers the trade agreement to the punishment while the lobby's preferences are the reverse. These details do not affect the main results and so are postponed to Section 6.1.

Here I focus on the key issue of the conditions under which the legislature decides to adhere to the trade agreement instead of violating it and triggering a punishment sequence. The trade agreement is broken when the legislature chooses a tariff that is higher than the trade agreement level,  $\tau^a$ . The level of the break tariff is chosen in the same manner as the trade war tariff, that is, according to Equation 3, but with the lobby's effort at the break stage  $e_b$  replacing its effort at the punishment stage,  $e_p$ .

The legislature will, however, only choose the break tariff over the trade agreement tariff if its continuation value from breaking the agreement is higher than the continuation value it receives from abiding by the agreement. The incentive constraint for the median legislator is essentially a condition on the trade agreement tariffs  $\tau^a$ . It can be written as

$$W_{ML}(\gamma(e_b), \tau^a) + \delta_{ML} V_{ML}^A \geq W_{ML}(\gamma(e_b), \tau^b(e_b), \tau^{*a}) + \delta_{ML} V_{ML}^P$$

where  $V_{ML}^A$  is the continuation value of the median legislator when it abides by the trade agreement  $V_{ML}^P$  is the continuation value when it defects and is punished.  $\delta_{ML}$  is the

---

<sup>17</sup>The most general condition that ensures that the lobby's second order condition holds is the following:  $\left| \frac{\partial \tau}{\partial \gamma} \frac{\partial^2 \gamma}{\partial e^2} \right| > \frac{\partial \pi}{\partial \tau} \left[ \frac{\partial \gamma}{\partial e} \right]^2 \frac{\frac{\partial^2 \pi}{\partial \tau^2}}{[ML's SOC]^2}$ . Note that to ensure concavity of the lobby's objective function, it's important that the decreasing returns to lobbying effort outweigh the direct impact of effort in increasing the weighting function. Also, if profits either increase too fast in tariffs or are too convex, the second order condition can be violated.

discount factor of the median legislator, while  $\delta_L$  will be used to represent the discount factor of the lobby and  $\delta_E$  that of the executive branch.

If the Nash reversion punishment lasts for  $T$  periods, then the only part of the continuation values that need be considered are the current period and the following  $T$  periods: after those  $T$  periods, the trade agreement will be in force so the continuation value will be the same from period  $T + 1$  on. Therefore we have<sup>18</sup>

$$W_{ML}(\gamma(e_b), \tau^a) + \frac{\delta_{ML} - \delta_{ML}^{T+1}}{1 - \delta_{ML}} W_{ML}(\gamma(e_b), \tau^a) \geq W_{ML}(\gamma(e_b), \tau^b(e_b), \tau^{*a}) + \frac{\delta_{ML} - \delta_{ML}^{T+1}}{1 - \delta_{ML}} W_{ML}(\gamma(e_b), \tau^{tw}). \quad (6)$$

Built into this condition is the legislature's applied tariff-setting behavior when the lobby's effort level is below the cutoff value  $\bar{e}(\tau^a)$  that leads the legislature to break the trade agreement. For any effort level weakly between  $e_a$  and  $\bar{e}(\tau^a)$ , the legislature chooses the applied tariff  $\tau^a$ .<sup>19</sup> If the effort level is below  $e_a$ , the legislature chooses the corresponding applied tariff, which is necessarily less than  $\tau^a$ . Because the lobby's net profits are highest at  $\tau^{tw}$ , when the lobby does not choose  $\bar{e}(\tau^a)$ , it will necessarily choose  $e_a$  and the applied tariff will be  $\tau^a$ .

The condition for the lobby is similar to Expression 6. Under the trade agreement, a break in the trade agreement, and punishment period, the lobby receives its profits at the chosen tariff level net of the effort level it exerts:

$$\pi(\tau^a) - e_a + \frac{\delta_L - \delta_L^{T+1}}{1 - \delta_L} [\pi(\tau^a) - e_a] \geq \pi(\tau^b(e_b)) - e_b + \frac{\delta_L - \delta_L^{T+1}}{1 - \delta_L} [\pi(\tau^{tw}) - e_{tw}]. \quad (7)$$

The next section explores the implications of these incentive constraints for the structure of the equilibrium trade agreement.

#### 4. Trade Agreement Structure

We can write the executives' joint problem as

$$\max_{\tau^a} \frac{W_E(\tau^a)}{1 - \delta_E} \quad \text{subject to} \quad (8)$$

<sup>18</sup>Note that  $\delta + \delta^2 + \dots + \delta^l = \sum_{k=1}^l \delta^k = \sum_{k=1}^{\infty} \delta^k - \sum_{k=l+1}^{\infty} \delta^k = \frac{\delta}{1-\delta} - \frac{\delta^{l+1}}{1-\delta} = \frac{\delta - \delta^{l+1}}{1-\delta}$ .

<sup>19</sup>Recall that  $e_a$  is the effort level at which the legislature chooses  $\tau^a$  as its optimal unilateral tariff.



$$\frac{\delta_{\text{ML}} - \delta_{\text{ML}}^{T+1}}{1 - \delta_{\text{ML}}} [W_{\text{ML}}(\gamma(e_b), \tau^a) - W_{\text{ML}}(\gamma(e_b), \tau^{tw})] \geq W_{\text{ML}}(\gamma(e_b), \tau^b(e_b), \tau^{*a}) - W_{\text{ML}}(\gamma(e_b), \tau^a) \quad (9)$$

and

$$e_b \geq \pi(\tau^b(e_b)) - \pi(\tau^a) + e_a + \frac{\delta_L - \delta_L^{T+1}}{1 - \delta_L} [\pi(\tau^{tw}) - e_{tw} - \pi(\tau^a) + e_a] \quad (10)$$

where Inequalities 9 and 10 are simple rearrangements of 6 and 7.

To understand how the executives optimally structure trade agreements subject to the given  $T$ -period Nash reversion punishment scheme, we must first examine the incentives of the lobbies and how the legislatures make decisions regarding breach of the trade agreement. The symmetric structure of the model permits restriction of attention to the home country.

I will consider the economically interesting case in which, for a given  $\delta = (\delta_E, \delta_{\text{ML}}, \delta_L)$  and  $T$ , there exists a non-trivial trade agreement in the absence of lobbying, that is, one in which the lowest supportable cooperative tariffs are strictly lower than the trade-war (i.e. non-cooperative) level. Call the trade-agreement tariffs in the absence of lobbies  $\tau_{\text{NL}}^a$ . If  $\tau_{\text{NL}}^a = \tau^{tw}$ , the lobby has no incentive to be active and the extra constraint implied by the presence of the lobby does not bind.

When deciding whether to exert effort to derail a trade agreement, the lobby has a two-stage problem. First, for the given  $\tau^a$ ,  $\delta$  and  $T$ , it calculates the minimum  $e_b$  required to induce the legislature to break the trade agreement. Call this minimum effort level  $\bar{e}(\tau^a)$ . This minimum effort level induces the minimum tariff that will break the agreement,  $\tau^b(\bar{e})$ .<sup>20</sup>

The following equation—Expression 9 at equality—defines  $\bar{e}$ :

$$\frac{\delta_{\text{ML}} - \delta_{\text{ML}}^{T+1}}{1 - \delta_{\text{ML}}} [W_{\text{ML}}(\gamma(\bar{e}), \tau^a) - W_{\text{ML}}(\gamma(\bar{e}), \tau^{tw})] - [W_{\text{ML}}(\gamma(\bar{e}), \tau^b(\bar{e}), \tau^{*a}) - W_{\text{ML}}(\gamma(\bar{e}), \tau^a)] = 0 \quad (11)$$

---

<sup>20</sup>Because it is assumed that the trade agreement commitment takes the form of a tariff cap (i.e. weak binding), only tariffs strictly greater than  $\tau^a$  serve to break the agreement.

This calculation of precise indifference is possible because it is assumed here that the political process is certain—that is, all actors know precisely how lobbying effort affects the identity of the median legislator through  $\gamma(e_b)$ .

Given the effort level required to break the agreement, the lobby will compare its current and future payoffs from inducing a dispute  $\left( \pi(\tau^b(\bar{e})) - \bar{e} + \frac{\delta_L - \delta_L^{T+1}}{1 - \delta_L} [\pi(\tau^{tw}) - e_{tw}] \right)$  to the profit stream from the trade agreement  $\left( \pi(\tau^a) - e_a + \frac{\delta_L - \delta_L^{T+1}}{1 - \delta_L} [\pi(\tau^a) - e_a] \right)$ . With the appropriate substitutions and rearrangements, this is just Condition (10) evaluated at  $\bar{e}$ . If the latter is larger, the lobby chooses to lobby only for the trade agreement tariff and the agreement remains in force. On the other hand, if the former is larger, the lobby induces the most profitable possible break.<sup>21</sup>

Anticipating this decision-making process of the lobby, the executives maximize social welfare by choosing the lowest tariffs such that the trade agreement they negotiate remains in force. They raise tariffs to the point that makes the lobby indifferent between exerting effort  $\bar{e}(\tau^a)$  to break the trade agreement and  $e_a$  to receive the trade agreement tariff.<sup>22</sup> That is, they choose tariffs so that the following equation holds:

$$\bar{e}(\tau^a) - [\pi(\tau^b(\bar{e})) - \pi(\tau^a) + e_a] - \frac{\delta_L - \delta_L^{T+1}}{1 - \delta_L} [\pi(\tau^{tw}) - e_{tw} - \pi(\tau^a) + e_a] = 0 \quad (12)$$

This is simply the lobby's constraint evaluated at  $\bar{e}$  when the lobby is indifferent.<sup>23</sup>

To understand the dynamics governing the solution to this problem, begin by considering the legislature's constraint at equality, Equation 11. This traces out a function from the trade agreement tariff into the minimum effort level required to break the trade agreement. The relationship between the home tariff and  $\bar{e}$  is straightforward.

**Lemma 1.** *The minimum lobbying effort required to break the trade agreement ( $\bar{e}$ ) is increasing in the home trade agreement tariff  $\tau^a$ .*

<sup>21</sup>Recall that the lobby's myopic profits are maximized at  $e_{tw}$ . This implies that if  $\bar{e} < e_{tw}$ , the lobby will break the agreement by exerting effort level  $e_{tw}$  since its net profits are higher than if it only exerted effort level  $\bar{e}$  in this case.

<sup>22</sup>Here I assume that the lobby chooses  $e_a$  when indifferent; if the opposite assumption were made, trade agreement tariffs would have to be raised by an additional  $\varepsilon$ .

<sup>23</sup>By construction, the legislative constraint will always be slack in equilibrium. The  $\bar{e}(\tau^a)$  schedule is calculated to make the median legislator indifferent between cooperating and initiating a dispute but then in equilibrium  $\tau^a$  is chosen so that the lobby does not break the agreement. When the lobby's effort level is less than  $\bar{e}(\tau^a)$ , the median legislator cannot prefer to break the agreement since her preferred tariff is lower when the lobby's effort is  $e_a$  than when it is  $\bar{e}(\tau^a)$ .

Proof: See the Appendix.

The intuition is as follows:  $\bar{e}$  must be at least as large as  $e_{tw}$  in equilibrium because the lobby's net profits are maximized at  $e_{tw}$ . If  $\bar{e} < e_{tw}$ , the lobby's constraint will not be satisfied since the lobby will exert effort level  $e_{tw}$ . Since the trade agreement tariff is weakly less than the trade war tariff, the median legislator's most preferred tariff at  $\bar{e}$ —that is,  $\tau^b(\bar{e})$ —must be higher than the trade agreement tariff. Raising  $\tau^a$  brings the trade agreement tariff closer to the legislature's ideal point, requiring the lobby to pay more to make the legislature willing to break the agreement.

The relationship between the foreign trade agreement tariff and  $\bar{e}$  is the opposite. This occurs because raising  $\tau^{*a}$  makes the agreement less attractive to the legislature and therefore requires less effort from the lobby to break.

**Lemma 2.** *The minimum lobbying effort required to break the trade agreement ( $\bar{e}$ ) is decreasing in the foreign trade agreement tariff  $\tau^{*a}$ .*

Proof: See the Appendix.

When the trade agreement is symmetric,  $\tau^a = \tau^{*a}$ . In this case,  $\bar{e}$  is concave in the trade agreement tariffs since the legislature's optimum in terms of  $\tau^a$  is at  $\tau^{tw}$  while its optimum in terms of  $\tau^{*a}$  is at zero.

The concavity of this  $\bar{e}(\tau^a)$  function implies that there may not be a truly interior solution to the executives' problem. Of course whenever the solution to the problem in the absence of lobbies cannot be satisfied for any  $\tau^a < \tau^{tw}$ , then the solution to (8) will also be  $\tau^{tw}$ . It may also be the case that there is a solution  $\tau^a < \tau^{tw}$  in the absence of the lobbies but that the lobbying constraint cannot be satisfied at any value other than the trade war tariff. The lobbying constraint will, however, always be satisfied at  $\tau^a = \tau^{tw}$  because there  $\pi(\tau^{tw}) - e_{tw} - \pi(\tau^a) - e_a = 0$ . Most of the results of this paper do not apply to this kind of solution, but it always exists and so a solution to the problem is guaranteed.

To see when an equilibrium of interest exists, recall that we need  $\bar{e} \geq e_{tw}$  in order for the lobby's constraint to be satisfied for  $\tau^a$  strictly less than  $\tau^{tw}$ . Even though it may appear at first sight that the constraint could be satisfied at a  $\tau^a$  for which  $\bar{e} < e_{tw}$ , in fact the lobby would choose the higher level of effort  $e_{tw}$  at which its net profits are maximized, breaking this incentive constraint.

If there does exist  $\tau^a < \tau^{tw}$  for which  $\bar{e} \geq e_{tw}$ , there *may* be another solution. What is required is that  $\bar{e}(\tau^a)$  does not begin to decrease too quickly before it can satisfy the lobby's constraint. The more easily the lobby can exert influence, the harder it is to satisfy this constraint: this causes  $\bar{e}$  to rise slowly with tariffs and keeps the price of a break low in comparison to the profits. It's quite intuitive that it is exactly when import-competing lobbies are strong that there may be no incentive compatible trade agreement that features positive levels of cooperation. It is not surprising that there are significant constraints on the existence of non-trivial trade agreements given that we observe many country-pairs and goods that are not covered by trade agreements.

Whether an interesting solution of the type we go on to examine in the next two sections exists or not, as long as there is a non-trivial trade agreement in the absence of lobbies, a trade agreement always exists and has the same form.

**Result 1.** *In the case of political certainty, the equilibrium trade agreement induces zero lobbying effort and is never subject to dispute. The executives choose the minimum tariff level that induces the lobby to choose  $e_b = 0$ .*

At the equilibrium tariffs, the lobby's constraint binds, while the legislature's does not. The amount of effort the lobby would have to exert to provoke a dispute, however, is derived from the legislature's constraint. This cost is then used in the lobby's constraint to calculate the lowest tariff level that will induce the lobby to disengage (that is, choose  $e_a$  over  $\bar{e}(\tau^a)$ ) and therefore make the median legislator's constraint slack and induce *her* to choose the internationally-agreed-upon  $\tau^a$  over  $\tau^b$  and the implied dispute.<sup>24</sup>

Although in this simple model we do not see disputes in equilibrium, the lobby's out-of-equilibrium incentives to exert effort to provoke a dispute are essential in determining the tariff-setting behavior of the executives.

---

<sup>24</sup>The results would be altered in magnitude but not in spirit by assuming that the trade agreement tariffs are strong bindings instead of weak bindings. The most important implication is that there would be zero lobbying effort in equilibrium as the lobby would not need to put forth effort to bid protection levels up to the tariff cap. There would still be no trade disputes in equilibrium.

#### 4.1. Trade Agreement Properties

Following Result 1, we know that the lobby first uses Expression 9 at equality to determine  $\bar{e}(\tau^a)$ : that is, it determines how much effort it has to exert for the given  $\tau^a$  in order to induce the legislature to choose noncooperation. This it accomplishes using Condition 11 above.

With  $\bar{e}(\tau^a)$  determined, the executives use Expression 12 at equality to determine the required  $\tau^a$ :<sup>25</sup> that is, the trade agreement tariff that is just high enough to induce the lobby to disengage at the break stage, causing the trade agreement tariff to remain in place in equilibrium.

Although one cannot arrive at explicit expressions for the solution functions  $\bar{e}(\cdot)$  and  $\tau^a(\cdot)$  without imposing further assumptions, significant intuition can be derived implicitly. An overview of the results will be provided here, while the mathematical details are in the Appendix. It's important to keep in mind that these results apply to solutions that are truly interior in the sense that the lobby has been disengaged by making it too costly to exert effort.

We begin with the comparative static question of how changes in the patience level of the lobby affect the equilibrium trade agreement tariffs.

**Corollary 1.** *As the lobby becomes more patient ( $\delta_L$  increases), the trade agreement tariff also increases, ceteris paribus.*

Proof: See the Appendix.

When the lobby becomes more patient, the equilibrium trade agreement tariff must be raised because the lobby now places relatively less weight on the lower net profits it gains during the break period relative to the benefits it attains during the trade war in future periods. The lobby's incentives to exert effort must be reduced by increasing the trade agreement tariff, thus reducing the profit gap between the trade war and the trade agreement.

A change in  $\delta_L$  might reflect a change in firms' planning horizons, or even their operational horizons—although it is not entirely clear in which direction this might work for firms who are facing extinction without sufficient protection. The lobby's

---

<sup>25</sup>There are analogous expression for  $\tau^{*a}$  throughout that can be ignored by symmetry.

patience level might also change with a change in the administrative leadership of the lobby, or as a reduced form for changes in risk aversion in a model with political uncertainty—a more risk-averse lobby would effectively weigh the future, uncertain gains less relative to the current, certain cost.

Turning to the patience of the median legislator, we start with the effect on the minimum lobbying effort level.

**Corollary 2.** *As the median legislator becomes more patient ( $\delta_{ML}$  increases), the minimum lobbying effort ( $\bar{e}$ ) required to break the trade agreement increases ceteris paribus.*

Proof: See the Appendix.

For any given level of effort, a more patient median legislator weighs the future punishment for deviating more heavily relative to the gain from the cheater's payoff in the current period. The lobby must compensate by putting forth more effort in the current period to bend the median legislator's preferences toward higher tariffs.

What does an increase in  $\delta_{ML}$ , leading to an increase in  $\bar{e}$ , imply for the optimal trade agreement tariff? The math is in the Appendix, but the intuition is straightforward.

**Corollary 3.** *As the median legislator becomes more patient ( $\delta_{ML}$  increases), the trade agreement tariff decreases ceteris paribus.*

Proof: See the Appendix.

This result contrasts with Corollary 1. When the median legislator becomes more patient, the executives are able to decrease the trade agreement tariff *because* the cutoff lobbying expenditure increases. This is because the lobby must now pay more to convince the legislature to choose short-run gains in the face of future punishment, so a wider profit gap between the trade war and trade agreement tariffs is consistent with disengaging the lobby.

Here the result comes through the legislature's indifference condition instead of directly from the lobby's indifference condition, but the intuition is the same: the trade agreement tariff is determined as whatever it takes to quell the lobby's willingness to exert effort to break the agreement.

The median legislator's patience level will increase with any change that makes her less susceptible to challenges from incumbents and therefore more likely to remain in office into the future. Changes to electoral rules, the strength of her party and similar political environment variables are influential here. Also influencing  $\delta_{ML}$  are electoral timing issues and individual decisions about seeking re-election.

Let's turn to another variable that impacts the equilibrium trade agreement in important ways: the weight the median legislator places on the profits of the import-competing sector. This political weighting function,  $\gamma(e)$ , is endogenous to many of the decisions underpinning the equilibrium, but here we examine the effect of an exogenous change in  $\gamma$ . First, on the cutoff effort level:

**Corollary 4.** *Exogenous positive shifts in the political weighting function  $\gamma(e)$  reduce the minimum lobbying effort ( $\bar{e}$ ) required to break the trade agreement, ceteris paribus.*

Proof: See the Appendix.

In accordance with intuition, if there is a shift in the political weighting function so that the legislature weights the profits of the import-competing sector more heavily for a given amount of lobbying effort, the lobby will have to exert less effort in order to induce a trade disruption.

This translates in a straightforward way to an impact on the trade agreement tariff.

**Corollary 5.** *Exogenous positive shifts in the political weighting function  $\gamma(e)$  lead to higher trade agreement tariffs, ceteris paribus.*

Proof: See the Appendix.

This makes sense given that an upward shift in the political weighting function in effect means that the lobby becomes more powerful, that is, it has a larger impact on the median legislator for a given level of effort. This is why the minimum effort level required to break the trade agreement is reduced, and therefore why the trade agreement tariff must be increased: when the lobby has to pay less to break the agreement for any given tariff level, the agreement must be made more agreeable to the lobby.

Examples of phenomena that would shift  $\gamma(\cdot)$  abound: the lobby becoming more effectively organized, a national news story that makes the industry more sympathetic in the eyes of voters, or the appointment of an individual who is particularly supportive

to a key leadership role in the legislature would all shift the political weighting function upward.

## 5. Optimal Dispute Resolution

In an environment without lobbying, ? show that social welfare increases (that is, trade-agreement tariffs can be reduced) as punishments are made stronger. This can be seen here if we restrict attention to the legislature's constraint:

$$\frac{\delta_{ML} - \delta_{ML}^{T+1}}{1 - \delta_{ML}} [W_{ML}(\gamma(e_b), \tau^a) - W_{ML}(\gamma(e_b), \tau^{tw})] \geq W_{ML}(\gamma(e_b), \tau^b(e_b), \tau^{*a}) - W_{ML}(\gamma(e_b), \tau^a)$$

This constraint is made less binding as  $T$  increases—that is, as we increase the number of periods of punishment. The intuition is straightforward: the per-period punishment is felt for more periods as the one period of gain from defecting remains the same. Thus larger deviation payoffs remain consistent with equilibrium cooperation as  $T$  increases.

**Lemma 3.** *The slackness of the legislative constraint is increasing in  $T$ .*

This is why the standard environment with no lobby gives no model-based prediction about the optimal length of punishment. Longer is better, although there are renegotiation constraints that must be taken into account that are typically outside of the model as well as other concerns.

The lobby's constraint

$$e_b \geq \pi(\tau^b(e_b)) - \pi(\tau^a) + e_a + \frac{\delta_L - \delta_L^{T+1}}{1 - \delta_L} [\pi(\tau^{tw}) - e_{tw} - \pi(\tau^a) + e_a]$$

works in the opposite direction in relation to  $T$ . Here, the lobby benefits in each dispute period, and so the total profit from a dispute is increasing in  $T$ . Thus we have

**Lemma 4.** *The slackness of the lobbying constraint is decreasing in  $T$ .*

Although the interaction of the impact of the length of the punishment on these two constraints is quite nuanced, in many cases, adding the lobbying constraint provides a prediction for the optimal  $T$  within this class of  $T$ -length Nash-reversion punishments.



As the executives choose the smallest  $\tau^a$  that makes the lobby indifferent at  $\bar{e}(\tau^a)$ , we must analyze the lobby's constraint evaluated at  $\bar{e}(\tau^a)$  (Expression 12) to determine the optimal length of punishment  $T$ . Obtaining the derivative of  $\bar{e}(\tau^a)$  from Equation 11 via the Implicit Function Theorem, the derivative of the lobby's constraint with respect to  $T$  is

$$\left(1 - \frac{d\pi}{d\bar{e}}\right) \frac{-\frac{\delta_{ML}^{T+1} \ln \delta_{ML}}{1-\delta_{ML}} [W_{ML}(\gamma(\bar{e}), \tau^a) - W_{ML}(\gamma(\bar{e}), \tau^{tw})]}{\frac{\partial \gamma}{\partial e} [\pi(\tau^b(\bar{e})) - \pi(\tau^a)] + \frac{\delta_{ML} - \delta_{ML}^{T+1}}{1-\delta_{ML}} \frac{\partial \gamma}{\partial e} [\pi(\tau^{tw}) - \pi(\tau^a)]} + \frac{\delta_L^{T+1} \ln \delta_L}{1-\delta_L} [\pi(\tau^{tw}) - e_{tw} - \pi(\tau^a) + e_a] \quad (13)$$

If this expression is negative for all  $T$ , the lobby's constraint is most slack at  $T = 0$ . The optimal punishment length cannot be zero, however, because the median legislator's constraint cannot be satisfied with a punishment period of length zero. In this case, which occurs only when the lobby is extraordinarily strong relative to the legislature, we must invoke an ad-hoc constraint on the minimum feasible length.

On the other hand, if this expression is positive for all  $T$ , the constraint is most slack as  $T$  approaches infinity and so we are in a case similar to that of the model without lobbying where a ad-hoc renegotiation constraint determines the upper bound on the punishment length. Here, the legislative constraint outweighs concerns about provoking lobbying effort. Perhaps of most interest are intermediate cases where the optimal  $T$  is interior—that is, the punishment length optimally balances the need to punish legislators for deviating with that of not rewarding lobbies too much for provoking a dispute.

The intuition is clearest if we examine the case of perfectly patient actors, that is, let  $\delta_L$  and  $\delta_{ML} \rightarrow 1$ . In essence, this removes the influence of the period of cheater's payoffs in which the interests of the legislature and the lobby are aligned (both do better in the defection stage) and exposes the differences between them in the dispute phase. In the limit, the derivative of the constraint with respect to  $T$  becomes

$$\left(1 - \frac{d\pi}{d\bar{e}}\right) \frac{W_{ML}(\gamma(\bar{e}), \tau^a) - W_{ML}(\gamma(\bar{e}), \tau^{tw})}{\frac{\partial \gamma}{\partial e} \{[\pi(\tau^b(\bar{e})) - \pi(\tau^a)] + T [\pi(\tau^{tw}) - \pi(\tau^a)]\}} - [\pi(\tau^{tw}) - e_{tw} - \pi(\tau^a) + e_a] \quad (14)$$

The proof of Corollary 3 shows that  $(1 - \frac{d\pi}{d\bar{e}})$  is positive.  $\bar{e}$  is determined so that

$W_{ML}(\gamma(\bar{e}), \tau^a) - W_{ML}(\gamma(\bar{e}), \tau^{tw})$  is always positive,<sup>26</sup> so the numerator of the first fraction is positive. The trade-agreement tariff is always lower than both the trade war tariff and the cheater's tariff ( $\tau^b(\bar{e})$ ) and  $\frac{\partial \gamma}{\partial e}$  is positive by Assumption 1, so the denominator is always positive. Note that the only influence of  $T$  on the entire expression is through this denominator, so the value of the expression is decreasing in  $T$ .

The second term, the lobby's gain from a break in the trade agreement, is always at least weakly positive since the trade agreement tariff will never be larger than  $\tau^{tw}$ . Note that whenever the lobby's constraint must be satisfied by choosing  $\tau^a$  such that  $\pi(\tau^{tw}) - e_{tw} - \pi(\tau^a) + e_a = 0$ , Expression 14 is always positive so that the optimal  $T$  is the largest possible value. Essentially, only the legislature's incentives are of concern in this case.

In the case of interest where the lobby potentially has an interest in breaking the agreement, the right-hand term is strictly positive. Here where we've taken  $\delta_L \rightarrow 1$ , the rate of change of the lobby's gain is constant.

Depending on the relative magnitudes, the overall expression may be positive for small  $T$  and then become negative, or it may be negative throughout. In the former case, the optimal interior  $T$  can be determined, while in the latter we must choose the shortest feasible  $T$ . The expression cannot be positive for all values of  $T$ , so it cannot be optimal to have arbitrarily long punishments when the players approach perfect patience.

**Result 2.** *Under Nash reversion punishments when both the legislature and lobby are perfectly patient, the optimal punishment scheme precisely balances the future incentives of the lobby and legislature. It always lasts a finite number of periods and may be of some minimum feasible length if the influence of lobbying on legislative preferences is extraordinarily strong (i.e.  $\frac{\partial \gamma}{\partial e}$  is sufficiently high).*

The key intuition for distinguishing between the situations described in Result 2 comes from examining the properties of the political process. If  $\frac{\partial \gamma}{\partial e}$  is moderate, the positive term in Expression 14 is more likely to dominate in the beginning and lead to an interior value for the optimal  $T$ , whereas extremely large values for  $\frac{\partial \gamma}{\partial e}$  make it more likely that the boundary case occurs. For a given effort level, this derivative

---

<sup>26</sup>See the discussion in the proof of Corollary 2 for a full treatment.

will be smaller when the lobby is less influential; that is, when a marginal increase in  $e$  creates a smaller increase in the legislature's preferences. Thus when the lobby is less powerful ( $\frac{\partial \gamma}{\partial e}$  is smaller), longer punishments are desirable. If the lobby is very influential, the same length of punishment will have a larger impact on the legislature's decisions (the impact on the gain accruing to the lobby does not change). This tips the balance in favor of shorter punishments.

This intuition generalizes for all  $(\delta_{ML}, \delta_L)$  as in Expression (13). Here the second-order condition is more complicated and can be positive if  $\frac{\partial \gamma}{\partial e}$  is very small. That is, if the lobby has very little influence in the legislature, it is conceivable that welfare will be maximized by making  $T$  arbitrarily large (subject, of course, to other concerns about long punishments).

**Result 3.** *Under Nash reversion punishments, if non-trivial cooperation is possible in the presence of a lobby, the optimal punishment scheme is finite when the influence of lobbying on legislative preferences is sufficiently strong ( $\frac{\partial \gamma}{\partial e}$  is sufficiently high).*

This helps to complete the comparison to the standard repeated-game model without lobbying. There, grim-trigger (i.e. infinite-period) punishments are most helpful for enforcing cooperation (cfr. ?'s Proposition 4). I have shown here that the addition of lobbies makes shorter punishments optimal in many cases. This is because long punishments incentivize the lobby to exert more effort to break trade agreements.

However, the model with no lobbies and one with very strong lobbies can be seen as two ends of a spectrum parameterized by the strength of the lobby. The optimal punishment will lengthen as the political influence of the lobby wanes and the desire to discipline the legislature becomes more important relative to the need to de-motivate the lobby.

## 6. Repeated Game Enforcement

### 6.1. Nash Reversion Punishments

I now turn to a full description of the strategies and incentives for the equilibrium that has been analyzed in the preceding sections—that is, when the punishment is Nash reversion for a limited number of periods. Again, this is a symmetric equilibrium, so

I describe strategies for the home country only; similar conditions hold for the foreign country.

In the first period of the game, in any period following a period when the trade agreement has been adhered to, after the successful completion of a punishment, or after a violation of the punishment by the lobby, the lobby chooses  $e_b = 0$  and the legislature chooses to abide by the agreement. In the applied tariff stage, the lobby chooses  $e_a$  and the legislature chooses  $\tau^a$  as the applied tariff.

In any period  $t$  in which a violation of the agreement occurred  $j+1$  periods previous for  $j \in [0, T-1]$  with Nash reversion punishments initiated  $j < T$  periods previous and followed in every period until  $t$ , the lobby chooses  $e_p \geq e_{tw}$  and the legislature chooses a tariff at least as large as its unilateral best response given  $e_p$ .

In any non-cooperative period in which the legislature does not follow its prescribed action, a new punishment phase is started. That is, the strategies prescribe that the lobby chooses  $e_p \geq e_{tw}$  and the legislature chooses a tariff at least as large as its unilateral best response given  $e_p$  for  $T$  periods.

In any non-cooperative period in which the lobby does not follow its prescribed action, a new cooperative phase is started. That is, the lobby chooses  $e_b = 0$  and the legislature chooses to abide by the agreement. In the applied tariff stage, the lobby chooses  $e_a$  and the legislature chooses  $\tau^a$  as the applied tariff.

Having fully described the strategies accompanying this punishment scheme, it must be shown that they constitute a subgame perfect equilibrium equilibrium, i.e. that each player's strategy constitutes a best response in every subgame. Recall that payoffs are independent of the trading partner's tariff level, so while checking incentives in the home country it can be assumed that players in foreign play their equilibrium strategies.

Section 3 establishes that the cooperative-phase behavior is incentive compatible for both the lobby and legislature given T-period Nash reversion punishment. Thus here we must show that it is incentive compatible to play the Nash-reversion punishments given the rest of the scheme. This will be accomplished in two phases: first by showing that the punishments are incentive compatible, and then by demonstrating the same for the punishments for deviating from the punishments.

First examine the incentives for the legislature to adhere to the punishment. La-

being the best deviation tariff as  $\tau^D$ , the incentive constraint in the first period of a punishment phase is

$$\begin{aligned} W(\gamma(e_{tw}), \tau^{tw}) + \frac{\delta_{ML} - \delta_{ML}^T}{1 - \delta_{ML}} W(\gamma(e_{tw}), \tau^{tw}) + \delta_{ML}^T W(\gamma(e_a), \tau^a) \geq \\ W(\gamma(e_{tw}), \tau^D, \tau^{*tw}) + \frac{\delta_{ML} - \delta_{ML}^T}{1 - \delta_{ML}} W(\gamma(e_{tw}), \tau^{tw}) + \delta_{ML}^T W(\gamma(e_{tw}), \tau^{tw}) \end{aligned} \quad (15)$$

In order to ensure that this condition holds, I assume that the median legislator prefers the trade agreement outcome to the trade war outcome:

**Assumption 3.**  $W(\gamma(e_a), \tau^a) \geq W(\gamma(e_{tw}), \tau^{tw})$

If Assumption 3 is violated, both the lobby and the legislature prefer the punishment and it will not be possible to enforce cooperation. Since  $\tau^{tw}$  is the legislature's one-shot best response to  $e_{tw}$ , there is no profitable deviation for the legislature given that it has no interest in creating a delay in completing the punishment. By the same logic, the legislature has no incentive to deviate in later portions of the punishment, as re-starting the punishment then would create a longer delay in returning to cooperation.

The lobby's condition for adhering to the punishment has the same structure:

$$\begin{aligned} \pi(\tau^{tw}) - e_{tw} + \frac{\delta_L - \delta_L^T}{1 - \delta_L} [\pi(\tau^{tw}) - e_{tw}] + \delta_L^T [\pi(\tau^a) - e_a] \geq \\ \pi(\tau^R(e_D)) - e_D + \frac{\delta_L - \delta_L^{T+1}}{1 - \delta_L} [\pi(\tau^a) - e_a] \end{aligned} \quad (16)$$

Since net profits are maximized at  $(e_{tw}, \tau^{tw})$ , this condition always holds. Any current period deviation creates a loss, and this only serves to create additional losses in the future. The size of the loss is the same in each period so the total loss grows with the number of periods in which the trade war tariff is replaced by the trade agreement tariff. This constraint is therefore tightest in the last period of the punishment but clearly holds for any  $T$  and at all points within a  $T$ -length punishment

There is a second type of subgame that must be checked: those in which one of the players has deviated from the prescribed punishments. If the legislature deviates from the punishment, that is, chooses a tariff that is less than its best response to  $e_{tw}$ , it must be incentive compatible for the lobby to exert  $e_{tw}$  in the next period to restart the punishment. The structure of incentives is exactly the same as in Expression 16.

Since it has been shown that Expression 16 holds for all  $j \in [1, T]$ , the abiding by the punishment is incentive compatible for the lobby.

Similarly, the legislature's constraint for abiding by the punishment is almost identical to Expression 15. Only the trade war effort level in the current period payoffs should be replaced by the deviation-from-the-punishment effort level. The logic is the same: the legislature can do no better than its instruction to best respond in the current period, and it would not want to deviate and take a loss in the present period because this only causes future losses. This condition also holds for all  $j \in [1, T]$  so incentive compatibility for the legislature is ensured. Thus the posited equilibrium supported by  $T$ -length Nash reversion punishments is subgame perfect.

## 6.2. Alternative Punishments

The above-explored symmetric Nash-reversion punishments are not the only possible punishments. This section explores an asymmetric punishment scheme in which welfare is reduced for both the legislature and the lobby in the defecting country. Although this type of punishment is hard to find in practice, I show here that in some cases it allows for increased cooperation in the form of lower trade agreement tariffs.

In this scheme, instead of  $T$  periods of Nash reversion, the legislature in the defecting country is required to apply a zero tariff for  $T$  periods, with an accompanying effort level of zero by the lobby. To keep the analysis simple, the non-defecting country's strategies are kept the same as in the Nash reversion case.

In order to compare the trade agreement tariffs under the two punishment schemes, I assume that both are interior in the sense that they are strictly less than  $\tau^{tw}$  themselves. Compared to the legislature's constraint under Nash reversion (Expression 11), the only change induced by the alternative punishment scheme is a reduction in the punishment tariff from  $\tau^{tw}$  to zero. This term in the constraint changes from  $W_{ML}(\gamma(\bar{e}), \tau^{tw}, \tau^{*tw})$  to  $W_{ML}(\gamma(\bar{e}), 0, \tau^{*tw})$ .

This results in an upward shift of the  $(\tau^a, \bar{e}(\tau^a))$  function. Intuitively, the punishment becomes harsher for the legislature, so the lobby has to exert more effort to achieve a break at any given level of  $\tau^a$ . We know that the punishment becomes harsher because  $\bar{e}$  must be at least as large as  $e_{tw}$  in equilibrium for the lobbying con-

straint to hold, implying that the legislature's welfare is maximized at some tariff at least as large as  $\tau^{tw}$ . Welfare therefore decreases as the tariff is reduced from  $\tau^{tw}$ , leading to the result that lobbying effort must increase for a given  $\tau^a$  when the tariff is reduced from  $\tau^{tw}$ .<sup>27</sup>

With this result in hand, we turn to the lobby's constraint. Here the changes are more complex. The constraint under symmetric  $T$ -period Nash reversion (Expression 12) can be written as

$$0 \geq [\pi(\tau^b(\bar{e})) - \bar{e}] - \pi(\tau^a) + \frac{\delta_L - \delta_L^{T+1}}{1 - \delta_L} [\pi(\tau^{tw}) - e_{tw} - \pi(\tau^a)]$$

Using tick marks to denote quantities under the alternative punishment scheme, this becomes

$$0 \geq [\pi(\tau^b(\bar{e}')) - \bar{e}'] - \pi(\tau^{a'}) + \frac{\delta_L - \delta_L^{T+1}}{1 - \delta_L} [\pi(0) - \pi(\tau^{a'})] \quad (17)$$

The first difference to notice is in the final bracket: the future payoff difference goes from positive to negative. The lobby no longer benefits from provoking a dispute. This serves to loosen the constraint.

Turning to the first bracketed term, it is shown above that  $\bar{e}' > \bar{e}$ . Since it must be the case that  $\bar{e} \geq e_{tw}$  where net profits are maximized, this implies that net profits decrease under the alternative punishment scheme. On initial impact, we hold  $\tau^a = \tau^{a'}$  so that the middle term is the same. Taken all together, these facts imply that the right side of Expression 17 becomes negative. Changing the punishment scheme thus creates slack in the lobby's constraint.

It is left to verify that this slack can be exploited to reduce  $\tau^a$ . It is clear that reducing  $\tau^a$  will tighten up the condition via the last two terms. What happens to net profits is not as straightforward since the  $(\tau^a, \bar{e})$  correspondence is concave.<sup>28</sup> If the reduction in the trade agreement tariff increases  $\bar{e}'$ , additional slack is created that

---

<sup>27</sup>To see this formally, fix  $\tau^a$  and examine the following comparative static exercise:  $\frac{d\bar{e}}{d\tau^p} = -\frac{\frac{\partial \Omega}{\partial \tau^p}}{\frac{\partial \Omega}{\partial \bar{e}}}$ . The numerator is  $-\frac{\delta_{ML} - \delta_{ML}^{T+1}}{1 - \delta_{ML}} \frac{\partial}{\partial \tau^p} W_{ML}(\gamma(\bar{e}), \tau^p, \tau^{*tw})$ . The argument in the preceding text shows that this is negative, while the denominator is shown to be negative in the proof of Lemma 1. Thus the expression as a whole is negative.

<sup>28</sup>It is easy to verify that the relationship remains concave under for this punishment scheme. There is the possibility that the results of Lemmas 1 and 2 could reverse because the denominator in each expression

allows further reduction in the trade agreement tariff. If instead  $\bar{e}^T$  is reduced, net profits will increase. This will limit the amount of reduction in  $\tau^{a'}$  that is consistent with satisfying the lobby's constraint. In either case,  $\tau^{a'}$  can be reduced below the level that is achieved under  $T$ -period Nash reversion punishments.

Thus this punishment scheme that is disliked by the lobby as well as the legislature can support lower trade agreement tariffs when it can be sustained. I now turn to the question of the incentive compatibility of the punishments.

Incentive compatibility depends on the full equilibrium strategies that are specified. It has just been shown that replacing the  $T$ -period Nash reversion punishments with a  $T$ -period punishment of zero tariffs and zero effort allows for a lower trade agreement tariff. Because the punishments for deviating from the trade agreement have a different structure, it's also necessary to find new punishments for deviating from the alternative punishments. It is desirable to design a scheme that allows for incentive compatibility of the trade agreement over the largest range of parameters.

Begin by examining the lobby's incentives for deviating from the punishment. Given that the legislature's action in a punishment period is to choose  $\tau^P = 0$  regardless of the lobby's choice and that the lobby is already exerting the lowest-cost effort level, the lobby cannot improve its payoff by deviating in the current period. The lobby's incentive constraint will hold if we specify that in the remaining periods of any punishment after a deviation by the lobby, the players play the original punishment actions. In effect, any deviation by the lobby is ignored.

Next, consider the legislature's incentives for deviating from the punishment. The legislature has a very real temptation to deviate to its best response tariff in the current punishment period. Note that with lobbying effort at zero, if one assumes that  $\gamma(0) = 1$ , this would be the standard optimal tariff for a large country. The legislature must be punished if it follows this deviation, and by standard repeated game logic, the punishment outcome should provide higher welfare to the lobby. The strongest such outcome

---

could be positive, but this requires  $\frac{\delta_{ML} - \delta_{ML}^{T+1}}{1 - \delta_{ML}} [\pi(\tau^a) - \pi(0)] > \pi(\tau^b) - \pi(\tau^a)$ . Since  $\pi(\tau^b) - \pi(\tau^a) > \pi(\tau^b) - e_b - \pi(\tau^a)$ , combining the two shows that in this case the lobby's constraint is violated. So Lemmas 1 and 2 hold as for the Nash reversion punishment.



is zero lobbying effort and a prohibitively-high tariff (call it  $\tau^P$ ). Denote the number of periods in this punishment as  $S_1$ , and note that this constraint is tightest during the first period of the punishment; later in the punishment the legislature receives a higher continuation value from following the prescribed punishment path because the trade agreement is restored more quickly. The relevant portion of the incentive constraint in the first period of the punishment is

$$W(\gamma(0), 0, \tau^{*tw}) + \frac{\delta_{ML} - \delta_{ML}^T}{1 - \delta_{ML}} W(\gamma(0), 0, \tau^{*tw}) + \frac{\delta_{ML}^T - \delta_{ML}^{S_1}}{1 - \delta_{ML}} W(\gamma(e_a), \tau^a) \geq \\ W(\gamma(0), \tau^D, \tau^{*tw}) + \frac{\delta_{ML} - \delta_{ML}^{S_1}}{1 - \delta_{ML}} W(\gamma(0), \tau^P, \tau^{*tw}) \quad (18)$$

Here  $S_1$  is taken to be at least as large as  $T$  and has to be calculated so that the condition holds. As it is quite likely that the gain from the optimal tariff is far outweighed by the loss from the prohibitive tariff for a non-politically-motivated legislature,  $S_1$  should not have to be much larger than  $T$ , if at all.

Next I turn to subgames in which one of the players has deviated from the prescribed punishment. If the legislature cheats on the punishment, the lobby has no incentive to deviate from enforcing legislature's punishment for deviating from the punishment  $(0, \tau^P)$  as, again, the lobby is already exerting the minimum effort. Thus ignoring any deviation by the lobby suffices to satisfy the lobby's incentives in this situation as well.

All that is left is to check that the legislature is happy to ignore any deviation by the lobby from the primary punishment. Consider a punishment like the one in Expression 18 except that it is of length  $S_2$ :

$$W(\gamma(e_{tw}), 0, \tau^{*tw}) + \frac{\delta_{ML} - \delta_{ML}^T}{1 - \delta_{ML}} W(\gamma(0), 0, \tau^{*tw}) + \frac{\delta_{ML}^T - \delta_{ML}^{S_2}}{1 - \delta_{ML}} W(\gamma(e_a), \tau^a) \geq \\ W(\gamma(e_{tw}), \tau^{tw}, \tau^{*tw}) + \frac{\delta_{ML} - \delta_{ML}^{S_2}}{1 - \delta_{ML}} W(\gamma(0), \tau^P, \tau^{*tw}) \quad (19)$$

For most economies,  $S_2$  will need to be larger than  $S_1$  in order for this constraint to hold. In any case, setting the punishment-for-deviating-from-the-punishment length to  $S = \max \{S_1, S_2\}$  will ensure that both constraints for the legislature hold. Then we have established the existence of an incentive compatible punishment scheme with zero tariffs and zero lobbying effort for all economies.

## 7. Conclusion

I have integrated a separation-of-powers policy-making structure with lobbying into a theory of recurrent trade agreements. This theory takes seriously the idea that the threat of renegotiation can undermine punishment when cooperation is meant to be enforced through repeated interaction alone. Assuming that countries can bind themselves to condition their negotiations on the state designation of a dispute settlement institution allows punishments to become incentive compatible.

I have shown here that, given no uncertainty about the outcome of the lobbying and political process, the executives maximize social welfare by choosing the lowest tariffs that make it unattractive for the lobbies to exert effort toward provoking a trade dispute. Although there are no disputes in equilibrium in this simple model, this extra constraint added by the lobby—apparently out-of-equilibrium—plays a key role in the determination of the optimal tariff levels and in the optimal dispute settlement procedure. While the constraint on the key repeated-game player, which here is the legislature, is loosened by increasing the punishment length, this new constraint due to the presence of lobbying becomes tighter as the punishment becomes more severe. This happens because the lobby *prefers* punishment periods in which tariffs, and thus its profits, are higher. It thus has increased incentive to exert effort as the punishment lengthens.

In a model with only the legislature, welfare increases with the punishment length. Here, this result only occurs if the lobby is sufficiently weak. As the lobby's political influence grows, the optimal punishment length becomes shorter—in the race between incentivizing the legislature and the lobby, the need to de-motivate the lobby begins to win. This suggests that a key consideration when designing the length of dispute settlement procedures is how to optimally balance the incentives of those capable of breaking trade agreements with the political forces who influence them, *given* the strength of that influence.

Future work is planned in at least two, related directions. In order for disputes to occur in equilibrium, I will add political uncertainty to the model as in ? (alternatively, asymmetric information could be introduced, or possibly both). The model will then be

able to address questions about the impact of political uncertainty on trade agreements and optimal dispute resolution mechanisms.

It will also be possible to explore whether accounting for the endogeneity of political pressure can explain the observed variation in the outcomes of dispute settlement cases (?) because, in this context, it becomes meaningful to ask when lobbies have the incentive to exert effort to perpetuate a dispute (hence removing the ad-hoc assumption imposed in the proof of Corollary 1 that the trade agreement tariff is always low enough to incentivize the lobby to exert effort during the trade war). Once political uncertainty has been added to the model, this is a completely natural extension that helps display the range and flexibility of the base model presented here.

## **8. Bibliography**

## Appendix A. Mathematical Appendix

### Proof of Lemma 1:

Labeling the left sides of Equations 11 and 12 as  $\Omega(\cdot)$  and  $\Pi(\cdot)$ , for notational convenience, these equations can be represented as<sup>29</sup>

$$\Omega(\bar{e}(\delta_{ML}, \gamma, \tau^a), \delta_{ML}, \gamma, \tau^a) = 0 \quad (\text{A.1})$$

$$\Pi(\tau^a(\delta_L, \delta_{ML}, \gamma), \bar{e}(\delta_{ML}, \gamma, \tau^a), \delta_L, \delta_{ML}, \gamma) = 0 \quad (\text{A.2})$$

By the Implicit Function Theorem:

$$\frac{d\bar{e}}{d\tau^a} = -\frac{\frac{\partial \Omega}{\partial \tau^a}}{\frac{\partial \Omega}{\partial \bar{e}}} = -\frac{\left[1 + \frac{\delta_{ML} - \delta_{ML}^{T+1}}{1 - \delta_{ML}}\right] \frac{\partial}{\partial \tau^a} W_{ML}(\gamma(\bar{e}), \tau^a)}{\frac{\delta_{ML} - \delta_{ML}^{T+1}}{1 - \delta_{ML}} \frac{\partial \gamma}{\partial \bar{e}} [\pi(\tau^a) - \pi(\tau^{tw})] - \frac{\partial \gamma}{\partial \bar{e}} [\pi(\tau^b(\bar{e})) - \pi(\tau^a)]} \quad (\text{A.3})$$

In order for the lobby's incentive constraint (Equation 12) to hold in equilibrium,  $\bar{e}$  must be at least as large as  $e_{tw}$ . Since the executives have no incentive to set the trade agreement tariff above the trade war tariff, this means that  $\tau^a \leq \tau^{tw} \leq \tau^b$ . Therefore  $\bar{e}$  will be set so that the median legislator's ideal point is (weakly) to the right of  $\tau^a$ , implying that the numerator is (weakly) positive.

Turning to the denominator,  $\gamma$  is assumed increasing in  $e$  so  $\frac{\partial \gamma}{\partial \bar{e}}$  is positive. Both profit differences are negative since  $\tau^a \leq \tau^{tw} \leq \tau^b$ . Therefore the denominator is negative.<sup>30</sup> Combined with the positive numerator and the leading negative sign, the expression is positive. ■

### Proof of Lemma 2:

By the Implicit Function Theorem:

$$\frac{d\bar{e}}{d\tau^{*a}} = -\frac{\frac{\partial \Omega}{\partial \tau^{*a}}}{\frac{\partial \Omega}{\partial \bar{e}}} = -\frac{\frac{\delta_{ML} - \delta_{ML}^{T+1}}{1 - \delta_{ML}} \frac{\partial}{\partial \tau^{*a}} W_{ML}(\gamma(\bar{e}), \tau^a)}{\frac{\delta_{ML} - \delta_{ML}^{T+1}}{1 - \delta_{ML}} \frac{\partial \gamma}{\partial \bar{e}} [\pi(\tau^a) - \pi(\tau^{tw})] - \frac{\partial \gamma}{\partial \bar{e}} [\pi(\tau^b(\bar{e})) - \pi(\tau^a)]}$$

<sup>29</sup>Note that all expressions also depend on the fundamentals of the welfare function— $D, Q_X, Q_Y$ —but these are suppressed for simplicity.

<sup>30</sup>Note that when  $\tau^a = \tau^{tw} = \tau^b$ , only the trivial trade agreement is possible and so this result and those that build upon it are not of interest.

The numerator is negative since the median legislator's welfare decreases in the foreign tariff (note that two other terms in the numerator cancel each other). The denominator is shown to be negative in the proof of Lemma 1. Combined with the negative numerator and the leading negative sign, the expression is negative. ■

**Proof of Corollary 1:**

By the Implicit Function Theorem:

$$\frac{d\tau^a}{d\delta_L} = - \frac{\frac{\partial \Pi}{\partial \delta_L}}{\frac{\partial \Pi}{\partial \tau^a}} = - \frac{\frac{1-(T+1)\delta_L^T + T\delta_L^{T+1}}{(1-\delta_L)^2} [\pi(\tau^{tw}) - e_{tw} - \pi(\tau^a) + e_a]}{\left(1 + \frac{\delta_L - \delta_L^{T+1}}{1-\delta_L}\right) \left[\frac{\partial \pi(\tau^a)}{\partial \tau^a} - \frac{\partial e_a}{\partial \tau^a}\right]} \quad (\text{A.4})$$

First I will show that  $\frac{1-(T+1)\delta_L^T + T\delta_L^{T+1}}{(1-\delta_L)^2}$  is positive. Focusing on the numerator and rearranging, we have

$$\begin{aligned} 1 - (T+1)\delta_L^T + T\delta_L^{T+1} &= (1 - \delta_L^T) - T\delta_L^T(1 - \delta_L) = (1 - \delta_L) \sum_{i=0}^{T-1} \delta_L^i - T\delta_L^T(1 - \delta_L) \\ &= (1 - \delta_L) \left[ \left( \sum_{i=0}^{T-1} \delta_L^i \right) - T\delta_L^T \right] = (1 - \delta_L) \left[ \sum_{i=0}^{T-1} \delta_L^i - \delta_L^T \right] > 0 \text{ for all } \delta_L < 1. \end{aligned}$$

Therefore  $\frac{1-(T+1)\delta_L^T + T\delta_L^{T+1}}{(1-\delta_L)^2}$  is positive.

The bracketed term is weakly positive since the trade agreement tariff is weakly smaller than the trade war tariff. In order for the results of this section to be interesting, it must be that  $\tau^a < \tau^{tw}$  so that the bracketed term is strictly positive for equilibria of interest.

Looking at the denominator, the discounting term is positive, so the term in parentheses is positive.  $\tau^a$  is weakly smaller than  $\tau^{tw}$  and net profits are increasing until  $\tau^{tw}$ , so the bracketed term is positive. As the product of two positive terms, the denominator is positive itself. Since both terms in the numerator have already been shown to be positive,  $\frac{d\tau^a}{d\delta_L}$  is positive. ■

**Proof of Corollary 2:**

By the Implicit Function Theorem:

$$\frac{d\bar{e}}{d\delta_{ML}} = -\frac{\frac{\partial \Omega}{\partial \delta_{ML}}}{\frac{\partial \Omega}{\partial \bar{e}}} = -\frac{\frac{1-(T+1)\delta_{ML}^T + T\delta_{ML}^{T+1}}{(1-\delta_{ML})^2} [W_{ML}(\gamma(\bar{e}), \tau^a) - W_{ML}(\gamma(\bar{e}), \tau^{tw})]}{\frac{\delta_{ML}-\delta_{ML}^{T+1}}{1-\delta_{ML}} \frac{\partial \gamma}{\partial \bar{e}} [\pi(\tau^a) - \pi(\tau^{tw})] - \frac{\partial \gamma}{\partial \bar{e}} [\pi(\tau^b(\bar{e})) - \pi(\tau^a)]} \quad (A.5)$$

I have shown in the proof of Corollary 1 that the first term in the numerator is positive. The bracketed term is positive because  $\bar{e}$  is always determined via Equation 11 so that  $W_{ML}(\gamma(\bar{e}), \tau^a) - W_{ML}(\gamma(\bar{e}), \tau^{tw})$  is positive: the trade-war tariff is the punishment relative to the trade agreement tariff. Therefore the numerator of the fraction is positive. The denominator is shown to be negative in the proof of Lemma 1. Therefore  $\frac{d\bar{e}}{d\delta_{ML}}$  is positive. ■

### Proof of Corollary 3:

Differentiating Equation A.2 with respect to  $\delta_{ML}$ , we have

$$\frac{\partial \Pi}{\partial \tau^a} \frac{d\tau^a}{d\delta_{ML}} + \frac{\partial \Pi}{\partial \bar{e}} \frac{d\bar{e}}{d\delta_{ML}} + \frac{\partial \Pi}{\partial \delta_{ML}} = 0$$

There is no direct effect of  $\delta_{ML}$  on this equation, so  $\frac{\partial \Pi}{\partial \delta_{ML}} = 0$ . Thus

$$\frac{d\tau^a}{d\delta_{ML}} = -\frac{\frac{\partial \Pi}{\partial \bar{e}} \frac{d\bar{e}}{d\delta_{ML}}}{\frac{\partial \Pi}{\partial \tau^a}} = -\frac{\left(1 - \frac{d\pi}{d\bar{e}}\right) \cdot \frac{d\bar{e}}{d\delta_{ML}}}{\left(1 + \frac{\delta_L - \delta_L^{T+1}}{1 - \delta_L}\right) \left[\frac{\partial \pi(\tau^a)}{\partial \tau^a} - \frac{\partial e_a}{\partial \tau^a}\right]} \quad (A.6)$$

The total effect of  $\bar{e}$  on  $\Pi$  is the negative of the lobby's FOC, that is

$$\frac{d}{d\bar{e}} [\bar{e} - \pi(\tau^b(\bar{e}))] = 1 - \frac{d\pi}{d\bar{e}} = -\left(\frac{d\pi}{d\bar{e}} - 1\right).$$

The lobby's FOC decreases to the right of  $e_{tw}$  since  $e_{tw}$  is the optimum ( $\frac{d\pi}{d\bar{e}} = 1$  at  $e = e_{tw}$ ).

Since we must have  $\bar{e} \geq e_{tw}$  in equilibrium in order for the lobby's constraint to bind, the effect of  $\bar{e}$  on  $\Pi$  is positive. In addition,  $\frac{d\bar{e}}{d\delta_{ML}}$  is positive by Corollary 2, so the numerator is positive.

By the same argument as in the proof of Corollary 1, the denominator is positive.

Since there is a leading negative sign,  $\frac{d\tau^a}{d\delta_{ML}}$  is negative. ■

### Proof of Corollary 4:

By the Implicit Function Theorem:

$$\frac{d\bar{e}}{d\gamma} = -\frac{\frac{\partial \Omega}{\partial \gamma}}{\frac{\partial \Omega}{\partial \bar{e}}} = -\frac{\frac{\delta_{ML}-\delta_{ML}^{T+1}}{1-\delta_{ML}}[\pi(\tau^a)-\pi(\tau^{tw})]-[\pi(\tau^b(e_b))-\pi(\tau^a)]}{\frac{\delta_{ML}-\delta_{ML}^{T+1}}{1-\delta_{ML}}\frac{\partial \gamma}{\partial \bar{e}}[\pi(\tau^a)-\pi(\tau^{tw})]-\frac{\partial \gamma}{\partial \bar{e}}[\pi(\tau^b(e_b))-\pi(\tau^a)]} \quad (A.7)$$

keeping in mind that the numerator is simplified by the envelope theorem. We can factor  $\frac{\partial \gamma}{\partial \bar{e}}$  out of the denominator and cancel the rest, leaving  $-\frac{1}{\frac{\partial \gamma}{\partial \bar{e}}} < 0$ . ■

### Proof of Corollary 5:

Differentiating the lobby's condition, Equation A.2 with respect to  $\gamma$ , we have

$$\frac{\partial \Pi}{\partial \tau^a} \frac{d\tau^a}{d\gamma} + \frac{\partial \Pi}{\partial \bar{e}} \frac{d\bar{e}}{d\gamma} + \frac{\partial \Pi}{\partial \gamma} = 0$$

Because  $\frac{\partial \Pi}{\partial \gamma} = -\frac{\delta_L-\delta_L^{T+1}}{1-\delta_L} \left[ \left( \frac{\partial \pi(\tau^{tw})}{\partial \tau^{tw}} - \frac{\partial e_{tw}}{\partial \tau^{tw}} \right) \frac{\partial \tau^{tw}}{\partial \gamma} \right]$ , we are looking for

$$\frac{d\tau^a}{d\gamma} = -\frac{\frac{\partial \Pi}{\partial \bar{e}} \frac{d\bar{e}}{d\gamma}}{\frac{\partial \Pi}{\partial \tau^a}} = -\frac{\left(1 - \frac{d\pi}{d\bar{e}}\right) \cdot \frac{d\bar{e}}{d\gamma} - \frac{\delta_L-\delta_L^{T+1}}{1-\delta_L} \left[ \left( \frac{\partial \pi(\tau^{tw})}{\partial \tau^{tw}} - \frac{\partial e_{tw}}{\partial \tau^{tw}} \right) \frac{\partial \tau^{tw}}{\partial \gamma} \right]}{\left(1 + \frac{\delta_L-\delta_L^{T+1}}{1-\delta_L}\right) \left[ \frac{\partial \pi(\tau^a)}{\partial \tau^a} - \frac{\partial e_a}{\partial \tau^a} \right]} \quad (A.8)$$

As shown in the proof of Corollary 4,  $\frac{d\bar{e}}{d\gamma}$  is negative, whereas the proof of Corollary 3 shows that  $\left(1 - \frac{d\pi}{d\bar{e}}\right)$  is positive. The trade war tariff is increasing in  $\gamma$ , as are net trade war profits. With the everywhere-positive discount term multiplying these two positive terms, we have another negative term because of the negative sign. Thus the numerator is negative.

The arguments given in the proof of Corollary 1 show that the denominator is positive. Therefore  $\frac{d\tau^a}{d\gamma}$  is positive when combined with the leading negative sign. ■