Gradualism in Free Trade Agreements:

A Theoretical Justification

Richard Chisik*

September, 1998

Revised December, 2001

Abstract

A notable feature of many recent trade agreements is the gradual, rather than immediate, reduction of trade barriers. In this paper we model trade liberalization as a cooperative relationship that evolves gradually in a non-cooperative environment. We show that specialization, capacity irreversibility and the development of trade-partner specific capital increase the benefit of continuing the liberalizing relationship and decrease, over time, the lowest obtainable self-enforcing tariff. By increasing the penalty of future defection, sunk costs ensure that the self-enforcing trading relationship starts slowly, but once in progress the level of cooperation continues to improve.

JEL Classification: F13, F15, C73.

Key Words: Trade Negotiations, Gradualism, Irreversibilities, Economic Integration, Dynamic Games.

---

* Department of Economics DM-309C, Florida International University, Miami, FL, 33199; E-mail: chisikr@fiu.edu. Phone: (305) 348-3286; Fax: (305) 348-1524.
I am grateful to Kyle Bagwell, Bruce Blonigen, Rick Bond, Barbara Craig, Ron Davies, Simon Evenett, Richard Harris, Kala Krishna, Devashish Mitra, Tom Prusa, Costas Syropoulos and to seminar participants at the University of Connecticut, The Penn State University, Florida International University, the University of Oregon, Rutgers University, the spring 1999 MWIEG Meetings, the 1999 CEA meetings, the 1999 SETI meetings, the 1999 SEA meetings, and the IEFS session at the 2000 ASSA meetings for their helpful comments. I would also like to thank Co-Editor Bob Staiger and an anonymous referee for many useful suggestions. I, alone, am responsible for any remaining errors.

I. Introduction

A notable feature of many trade agreements in the post World War II era is the gradual, rather than immediate, reduction of trade barriers. For example, European, Japanese and U.S. average tariffs have declined by over 90 percent since the first GATT round. These gradual reductions are partly attributable to deeper tariff cuts in successive GATT rounds, partly attributable to an increase in the types of trade barriers addressed by the GATT and partly attributable to an expansion in the number of sectors covered by the GATT. Furthermore, starting with the Kennedy round negotiations, new tariff reductions were given an explicit phase-in period. Many regional and bilateral agreements exhibit these same incremental reductions and well-detailed phase-in periods.[1]

Because of their smaller scale, bilateral trade agreements allow us to more clearly see the gradual increase in economic integration that may be generating these incremental reductions. For example, recent trade agreements with Canada, Israel and Mexico were preceded, if not initiated, by an unilateral reduction in U.S. tariffs levied against the prospective partner. As a result of this conventional unilateralism these countries increasingly oriented their exports toward the U.S. and were, therefore, more susceptible to a subsequent increase in aggressive unilateralism (increased use of anti-dumping duties, countervailing measures and safeguard actions) by the U.S. The free trade agreement essentially constituted a reciprocation of the U.S. tariff reductions along with a promise of further concessions by both parties. The U.S.-Israel free trade agreement provides a straightforward example. The agreement called for the removal of all tariffs and many other forms of protection over the course of a ten-year period beginning with implementation on September 1, 1985. The U.S. tariff reductions were quick, and relatively easy, with many Israeli goods already entering the U.S. duty free under the U.S. Generalized

[1] In Mercosur, for example, which was formed on March 26, 1991, tariffs levied by Argentina and Brazil were to decline by sixty-one percent by June 30, 1992 and by one-hundred percent by December 31, 1994. Paraguay and Uruguay were given an additional phase in year. In addition to the covered sectors, the treaty outlined plans for harmonization on intellectual property protection and foreign direct investment and found the possibility of future harmonization on anti-dumping duties and countervailing measures to be desirable. For more on Mercosur, see Rowat et al (1997).

System of Preferences.[2] The first order of the agreement was, therefore, essentially a reciprocating reduction in Israel's tariffs against U.S. imports. In response, the U.S. was to curb its use of aggressive unilateralism. Israel would then liberalize some of its licensing procedures. Furthermore, a nonbinding commitment to liberalize trade in services between the partners was part of the U.S.-Israel agreement.

As illustrated above, there is a similarity in the gradual steps that were taken in large multilateral agreements, such as the GATT, and in many regional and bilateral agreements. In the bilateral agreements, we also see an example of the mechanism that could lead to this gradualism process: Initial tariff reductions may lead to a growing interdependence between the trading partners.

In this paper, we develop a simple model that addresses these stylized facts. Trade liberalization emerges as a cooperative relationship that evolves gradually in a non-cooperative environment. Our point of departure is the explicit recognition of the non-stationary aspect of the trading environment. In particular, we show that specialization and the development of trade-partner specific capital increase the benefit of continuing the liberalizing relationship and decrease, over time, the lowest obtainable self-enforcing tariff.

The idea here is that each country must alter, or augment, its production process to realize the benefits of a trade agreement. These alterations may arise from increased output in an export sector, they may fit exports to the importing country's standards or they may involve network and sales infrastructure development in the importing country. Once the alteration cost is sunk, the value of continuing the liberalized trading relationship increases. For example, specialization in an export sector greatly reduces the elasticity of export supply and, in a non-cooperative equilibrium, greatly increases the optimal tariff that can be levied against the specializing country. If only one country makes an alteration, that country is exposed to opportunism and, foreseeing this possibility, production transformation and trade

---

[2] Israel's accession to the GATT, as part of the agreement, and its subsequent receipt of MFN status further reduced the number of Israeli exports subject to U.S. tariffs. Approximately eighty percent of Israeli exports to the U.S. and fifty percent of U.S. exports to Israel entered duty free on the agreement's inception. Naturally, the most sensitive sectors had the longest phase in periods. For more on the U.S.-Israel free trade agreement, see Rosen (1989).

liberalization are limited.[3] If both countries make alterations, then each proceed slowly, however, once it has been verified that the alteration costs are sunk, each party agrees to a greater degree of liberalization. By relaxing the incentive constraints, a larger degree of cooperation is sustainable. Put another way, by increasing the penalty of future defection, sunk costs ensure that the self-enforcing trading relationship starts slowly, but once in progress the level of cooperation continues to progress.

We consider dynamically optimal tariffs in a model of two symmetric economies. Our key assumption is the irreversibility of capacity accumulation in the export sector. Factor allocation in each period, therefore, is determined by the expected terms of trade and by the previous period allocations.

Our first result shows that if capacity reversibility is sufficiently slow, or costly, then, irrespective of the discount rate, free trade cannot be self-enforcing in the initial period of the trade agreement. A smaller tariff reduction is self-enforcing and in response to this tariff reduction additional factors begin to accumulate in the export sector. We, therefore, next characterize a set of equilibria where tariffs decline over time and, if governments are sufficiently patient, then the tariffs converge to the free trade cooperative outcome. An interesting feature of the model, here, is that, when capacity is sufficiently irreversible, tariff reductions must exhibit gradualism if they are reduced at all. In these equilibria, future liberalization depends on past successes and on the expectation of future gains.

We also show how the obtainable tariff outcomes are affected by changes in the extent of capacity irreversibility and by changes in our index of the appropriable gains from trade. First, we demonstrate that an increase in the extent of capacity irreversibility (i.e. if capacity reversibility becomes slower, or more costly) makes it more difficult to start a new trade agreement but makes it easier to sustain a well-established one. That is, an increase in capacity irreversibility raises the deviation

[3] Although tariff revenue was, historically, an important concern of trade policy makers, it appears to be of secondary importance, in industrialized countries, in the modern era. A natural response to this fact is to question the fear of opportunism as an explanation behind the development of recent trade agreements. (For some fascinating historical accounts of opportunism in trade relationships see McLaren, 1997). Whether the tariff levied against an increasingly specialized country is raised because of opportunism and tariff revenue concerns or is raised in response to domestic lobbyists demanding greater protection from growing import penetration is immaterial. The specializing country need only note that the tariff that prevails if the trade agreement fails is increasing in import

incentive in a new trading relationship but raises the cost of a trade war by a greater amount in a developed one. Second, we clarify that in the standard case, when capacity is instantly reversible, an increase in the gains from trade has an ambiguous effect on tariff outcomes. This result occurs because an increase in the gains from trade raises the incentive to deviate from the free trade agreement, but it also raises the cost of the future trade war that is generated by this deviation. Third, we establish that if the extent of irreversibility is sufficiently high, then an increase in the gains from trade decreases the cost of a future trade war as well. Hence, the lowest obtainable tariff must increase and it is seen that an increase in the gains from trade can perversely retard the gradualism process.

Finally, in characterizing the tariff path, we obtain a gradual counterpart to stationary folk theorem results: The speed of tariff liberalization is increasing in the discount factor. Put another way, if countries place a higher value on future payoffs, then the magnitude of tariff cuts is greater and the number of rounds of tariff reductions needed for free trade to obtain is reduced.

This paper is divided into six sections. The next section reviews the related literature. The third section develops the basic model and examines the stationary outcomes. The fourth section shows how gradualism occurs in a specific case. The fifth section provides general conditions under which gradualism can occur and examines the gradual tariff path. The sixth section concludes and suggests possibilities for further extensions. Proofs of the more technical propositions are available upon request from the author.

II. Related Literature

Johnson (1953) was the first to consider tariff retaliation and a trade-war equilibrium. Mayer (1981) used the Nash-equilibrium trade war as a threat point in a cooperative approach to characterizing trade agreements. The most famous result from Mayer's paper was that, because country size can determine the threat point, a large country may prefer a trade war over free trade with a small country. Put another way, the large country may need to be bribed to accept free trade.

---

penetration. In our model, as in many common frameworks, this is, in fact, the case.

In McLaren (1997), which is an important extension of Mayer's cooperative approach, factor allocation precedes a trade agreement. Because governments can give side payments, agents do not internalize the erosion in national bargaining power caused by their actions. If free trade is expected, then factors will accumulate in the export sector causing an increase in the optimal tariff that can be levied against this country. In this case, the resulting side payment in the trade agreement may be so large as to leave the country worse off under an optimistic expectation of free trade than under an expectation of a trade war.

Lapan (1988) was the first to formally recognize that the optimal tariff after production has occurred is greater than the ex-ante optimal tariff.[4] Internalizing this time inconsistency in tariff setting can lead to lower output levels and leave both countries worse off.

A potential limitation of analyzing trade agreements with a cooperative game theoretic approach is the lack of an international agency with true enforcement capability. Recognizing this fact, authors such as Dixit (1987) and Bagwell and Staiger (1990, 1997) began to look at trade agreements as self-enforcing outcomes in a repeated game framework.

Staiger (1995) explicitly considered gradual tariff reductions. In his paper, gradualism arises from the presence of rent-earning factors that are displaced by trade liberalization. Each successive round of trade liberalization displaces a small percentage of factors in the import competing industry. If these factors lose their rent-earning skill, which occurs with some exogenous probability, then they will have no rents to protect in subsequent liberalization rounds and further tariff cuts can occur. In a related vein, Furusawa and Lai (1998) showed that gradualism can arise if there are adjustment costs in moving

[4] Lapan's result is for national income maximizing tariffs. Staiger and Tabellini (1987) demonstrated a similar time inconsistency in setting politically optimal tariffs: Governments announce a low tariff to encourage workers to move out of the import-competing sector, but ex-post set a higher tariff to avoid adjustment costs.

workers in and out of the import-competing sector.[5]

In Devereux (1997), learning-by-doing augments the increased export-sector output afforded by liberalization to generate a reduction in production costs. These lowered costs, in turn, increase the benefit to consumers of further liberalization. Firms specialize immediately, therefore, gradualism arises only because the threat of a trade war becomes increasingly unfavorable to consumers.[6]

Bond and Park (forthcoming) also looked at gradualism arising from time-inconsistency in tariff setting by a large country. There are no capacity irreversibilities in their framework, therefore, the economic environment looks the same in every period. In contrast to ours, and all of the other gradualism frameworks, it is not the evolution of a state variable that drives the gradualism result. Gradualism arises instead because the small country desires to smooth consumption over time. The trade agreement may, therefore, stipulate higher tariffs in initial periods and lower tariffs with larger side payments from the small to the large country in later periods.

Our approach extends this literature in the following way. Unlike Staiger and Furusawa and Lai, who model further tariff reductions as arising from a reduced import competing sector, we, like Devereux, model gradualism as arising from economic integration and increasing interdependence between the trading partners. In our model, however, it is not consumers so much as producers that become more dependent on trade liberalization. It is the increasing export orientation of producers, here, that leads countries to lower tariffs in successive negotiating rounds. Finally, we analyze how changes in capacity

[5] Adjustment costs and the presence of rent earning factors in the import sector provide a useful approach to understanding the slow pace of trade liberalization, however, we believe that there are additional dynamic forces at work in the structure of trade agreements and we propose an alternative theory of gradualism that captures some of these effects. For example, after Mexico signed a trade liberalization agreement with the U.S. it quickly signed another with France. If the fear of initially displacing too many rent-earning factors was the main impetus behind the gradual tariff reductions written into NAFTA, then signing another trade agreement would not alleviate this factor displacement.

[6] Our model is concerned with irreversibilities, and does not pretend to capture Devereux's sophisticated change in the gains from trade. Our admittedly simplistic, and reduced form, index of this change, however, suggests that the Devereux effect may be mitigated when capacity is highly irreversible.

irreversibility, the gains from trade and the discount factor affect the obtainable tariff outcomes.[7]

Finally we wish to acknowledge the work of several researchers who showed that gradualism in unilateral trade liberalization may be national welfare maximizing for a country with domestic market imperfections. Mussa (1986) demonstrated the superiority of gradualism when adjustment costs are convex in the number of workers leaving the import-competing sector. Lapham (1995) finds a similar result in a dynamic monopolistic competition model with convex adjustment costs. Mehlum (1998) shows that gradual tariff reductions may improve welfare in a Ramsey type model encompassing Heckscher-Ohlin trade when a minimum wage is introduced (so that tariff reductions can cause transitional unemployment). Falvey and Kim (1992) discuss how gradualism may be justified when a government has not yet established an alternative means of generating revenue. They also note that large tariff reductions may not be considered credible by adjusting firms. We complement these normative justifications for unilateral gradualism in the presence of domestic distortions by analyzing a positive explanation for bilateral gradualism that arises from international strategic considerations.

III. Economic Environment

In this section, we consider an infinitely repeated tariff setting game between the governments of two production economies. The model is kept simple, not only to ensure tractability, but also to demonstrate gradualism arising solely from the presence of irreversible investments.

A. Assumptions on Tastes and Technology.

There are two symmetric economies. Each economy contains a government, a large number of firms and a large number (ℓ) of identical agents who are each endowed with one unit of labor. These agents sell their labor to the firms, and receive an equal share of the firms' profits and of the

---

[7] In a related application, Admati and Perry (1991) consider two agents who take turns making non-refundable contributions to a public project that is produced only if the sum of the contributions is greater than the cost. This framework generates a potential hold-up problem, so that agents contribute slowly, if at all. Our model is closer to what Admati and Perry refer to as the "standard formulation where the question is the scale of the public good that is provided." In their model it is all or none – although contributions are made incrementally, there is no gradualism in providing the public good. Furthermore, their incremental contribution result, and possible inefficiency result would not obtain in the "standard formulation." We exhibit gradualism in an international trade version of the "standard

government's tariff revenue. They spend this income on consumption of the firms' products. The strategic possibilities of the agents and the firms is limited by their large numbers and is, therefore, ignored in the set of subgame-pefect equilibria that we analyze below. Given the competitive behavior of the agents and the firms, each government chooses tariffs to maximize national welfare.

Each country is identified with a good in that only firms in country x (y) have the specific factor necessary for profitable production of good x (y). Gradualism, therefore, cannot arise, here, from employment rigidities in the import competing sector (there isn't one). In addition, there is a numeraire good (z). The numeraire good can rectify trade imbalances and its sector can costlessly and instantaneously contract to provide labor for a growing export sector.

The preferences of the identical agents, in each country, over consumption of the three goods can be represented by a quasilinear utility function. We can, therefore, restrict our analysis to the aggregate utility function. Letting countries be denoted by superscripts and goods by subscripts, for country i this utility function takes the following form:

$$U^i(D^i_x, D^i_y, D^i_z) = u(D^i_x) + u(D^i_y) + D^i_z. \qquad (1)$$

The sub-utility functions are quadratic: $u(D^i_j) = A \cdot D^i_j - (D^i_j)^2 / 2$. Using a prime to denote a partial derivative, market demand for good j in country i in period t is, therefore, given by:

$$D^i_{jt} = D^i_j(P^i_{jt}) = (u')^{-1}(P^i_{jt}) = A - P^i_{jt}; \qquad i \in \{x, y\}, \qquad j \in \{x, y\}. \qquad (2)$$

$$D^i_{zt} = w^i_t \cdot \ell + \sum_{j \in \{x,y\}} (\pi^i_{jt} + \tau^i_{jt} \cdot M^i_{jt} - P^i_{jt} \cdot D^i_{jt}). \qquad (3)$$

The wage in country i in period t is w^i_t and $\pi^i_{jt}, \tau^i_{jt}, M^i_{jt}, P^i_{jt}$ are, respectively, the profits from, the tariff on, the imports of, and the price of good j in country i in period t.

Good z is produced by a constant-returns-to-scale technology: $Q^i_{zt} = \ell^i_{zt}$, where Q^i_{zt} is the output of, and ℓ^i_{zt} is the labor input to, the numeraire sector in country i in period t. Because the labor supply is

---

formulation" because investments are not only sunk but also ongoing and temporarily irreversible.

assumed sufficiently large (to be made precise below), there is positive numeraire production in both countries and the wage is equal to the price of the numeraire good, which is normalized to one.

Goods x and y require the installation of **capacity** (Q^i_{jt}) before their production is viable. There is an increasing opportunity cost of adding capacity. This cost function is quadratic and its per-period marginal cost is given by:

$$C^i_{jt} = \begin{cases} B + Q^i_{jt} & \text{for} \quad i = j \\ A + Q^i_{jt} & \text{for} \quad i \neq j, \end{cases} \tag{4}$$

where, B < A. This capacity installation is not a one time expenditure but must be paid in every period. The minimum cost of producing good j in country i ≠ j is, therefore, greater than the maximum price at which demand is positive. Hence, equation (4) formalizes the assumption that only country x (y) can profitably produce good x (y) in the absence of a government subsidy.

To isolate the role of irreversible capacity investments, we assume that each industry's marginal production cost is zero for output less than or equal to capacity but is infinite for output greater than capacity.[8] The per-period production costs are then equal to the capacity costs and the world output of good x (y) is equal to country x's (y's) capacity: Q^x_{xt} (Q^y_{yt}) . The gain from international trade is, therefore, an increasing function of the magnitude A − B and we refer to θ = A − B as an index of the gains from trade.

Within a single time period, the capacity requirement is indistinguishable from a simple upward-sloping industry supply curve and allows for a straightforward analysis and comparison of per-period equilibrium values. The importance of the capacity requirement, however, is its effect between, or across, time periods. In particular, we assume that capacity is irreversible for N periods after the most recent capacity expansion. Equivalently, capacity cannot be reduced for N periods after the latest addition to capacity is made. This assumption implies that expanding output in any single time period observably

[8] Allowing for a positive production cost would mildly complicate the analysis without altering the results.

constrains output choices for the next N periods.[9] The **extent of irreversibility**, as measured by N, is shown below to influence the set of attainable trade agreement outcomes. This assumption is stated formally as:

$$\text{If } Q^i_{js} > Q^i_{j(s-1)}, \text{ then } Q^i_{j(s+t)} \geq Q^i_{js}; \quad \text{for } t \in \{1, \ldots, N\}, \quad i \in \{x, y\}, \quad j \in \{x, y\}. \tag{5}$$

From equations (4) and (5) it is evident that $C^i_{j(s+t)} \geq C^i_{js}$ for N periods after the latest capacity expansion. The idea here is that expanding an export sector requires the development of sector-specific human or physical capital. The labor that is devoted to creating this capital is no longer available for numeraire production. This cost, in terms of lost numeraire production, has to be paid whether or not the installed capacity is used.

There are several ways to motivate this capacity irreversibility assumption. The simplest assumes that firms have contractual obligations with their input suppliers arising, for example, from union contracts. These irreversibilities may also occur because of the need to develop and maintain networks and sales infrastructure in the importing country. Some of these expenses are sunk at the time of export expansion, however, many are also ongoing costs whose irreversibility again stems from explicit contracts (such as advertising, brand name and sales infrastructure maintenance) and implicit contracts (such as maintaining networks and political favor). Roberts and Tybout (1997) provide evidence that, for Colombian firms, these costs are an important component of the decision to enter an export market. Finally, these sunk costs may arise from an increase in capital expenditures (such as factory size and

[9] Our key assumption is that firms must take some observable action when they decide to reverse capacity and that this desired reversal is not immediate or costless. As it would have no effect on the outcome, we choose not to complicate the model by the addition of an additional variable that announces this capacity reversal decision and we endogenize it as in equation 5. In a related vein, it is also important that capacity is observable to the foreign government, for without this observable hostage gradualism would not occur. This point is related to Bagwell (1995) who showed how unobservability can eradicate a first-mover advantage. Maggi's (1999) extension of, and solution to, the Bagwell effect may apply here as well if the firms' capacity costs are private information, however, it would require that individual firms are willing to incur costly signaling to provide credibility to their government's commitment to a low announced tariff. Additionally, we rely on perfect observability of tariffs throughout our analysis. We believe that our results would still obtain if trade protection instruments are imperfectly observed as in Hungerford (1991) or Reizman (1991), however, the gradualism path may be non-monotonic. We leave the analysis of imperfect and incomplete information in our framework for further research.

machinery). Rather than paying the full cost at the time of installation, the capital cost is amortized over its lifespan, therefore, the per-period cost can be considered as loan repayment.

In addition to the firm's capacity choice, each government can choose to levy a per-unit import tariff. Given that country x exports x and imports y in any trading equilibrium, these tariffs relate the prices in the two countries as follows: $P^y_{xt} = P^x_{xt} + \tau^y_{xt}$; $P^x_{yt} = P^y_{yt} + \tau^x_{yt}$ which can be more succinctly expressed as $P^i_j = P^j_j + \tau^i_j$, where P^j_j is the price of good j in country j and P^i_j is the tariff-inclusive price paid by consumers in the importing country. From this point onward, when not necessary for clarity, time subscripts are omitted to unclutter notation.

The equality between world demand and supply for each good, $Q_j = D_j = D^x_j + D^y_j$, combined with equations (2) and the above pricing relationship implicitly defines the price of each good in each country as:

$$P^i_j = P^i_j(Q_j, \tau^i_j) = (2A - Q_j + \tau^i_j)/2; \quad P^j_j = P^j_j(Q_j, \tau^i_j) = (2A - Q_j - \tau^i_j)/2\,. \tag{6}$$

Substituting equation (6) into equation (2) yields each country's demand for good j as:

$$D^j_j = (Q_j + \tau^i_j)/2; \quad D^i_j = (Q_j - \tau^i_j)/2\,. \tag{7}$$

Equating P^j_j to C^j_j yields the unconstrained world supply of good j as:

$$Q_j = [2\theta - \tau^i_j]/3, \tag{8}$$

Equations (6) through (8) are derived under the assumption of non-prohibitive tariffs, which is later shown to be the case in all equilibria.

Tariffs are chosen by each government to maximize their respective country's welfare. Using equations (1 – 4) per-period welfare can be represented as the aggregate indirect utility function. It is the sum of consumer surplus, producer surplus, tariff revenue and the market value of the labor endowment:

$$V^i(\tau^i_j, \tau^j_i, Q_i, Q_j) = \int_{P^i_i(\tau^j_i)}^{A} D^i_i(P)dP + \int_{P^i_j(\tau^i_j)}^{A} D^i_j(P)dP + \int_{B}^{P^i_i(\tau^j_i)} Q_i(P)dP + \tau^i_j \cdot D^i_j(P^i_j(\tau^i_j)) + \ell\,. \tag{9}$$

Country i's imports of good $j \neq i$ equals their total demand for good j, therefore, the second-to-last term in equation (9) is tariff revenue. Letting $\delta \in [0, 1]$ represent the discount factor between periods, welfare in country i is then: $\sum_{t=0}^{\infty} \delta^t \cdot V^i(\tau^i_{jt}, \tau^j_{it}, Q_{it}, Q_{jt})$.

B. Timing.

In the beginning of each period firms and the government in each country observe the trade agreement. The agreement specifies the most cooperative self-enforcing tariff obtainable in that period. We assume that firms in each country know their respective government's tariff choice before their capacity decision is made. This implies the following timing of events.

First, tariffs are chosen simultaneously by governments in each country. The tariffs are initially revealed only to the firms in each country. Second, firms in each country simultaneously add capacity. The large number of firms implies that capacity is installed until its marginal cost equals the expected market price. Third, tariffs and capacities are revealed and production and consumption take place.

The above timing indicates that the public history in period t+1 is the sequence of tariffs and capacities chosen in each of the t previous periods. Formally, let $\tau^{it}_j = \{\tau^i_{j0,}, ..., \tau^i_{jt}\}$ and $Q^t_j = \{Q_{j0,}, ..., Q_{jt}\}$. The public history in period t+1 is $H^{t+1} = \{\tau^{xt}_y, \tau^{yt}_x, Q^t_x, Q^t_y\}$ and $H^0 = \{\emptyset\}$. Using this formalization, a strategically equivalent extensive form that is closer in spirit to more common, two player, games of almost perfect information is shown below.[10]

FIGURE 1 GOES APPROXIMATELY HERE

[10] Tariff negotiations are often analyzed as the outcome of a simultaneous move game between the governments of two endowment economies. The analysis can, therefore, draw on well-established theorems for analyzing such games of almost-perfect information. The difficulty here is that: (a.) there is production in the economy and (b.) that capacity is irreversible. Firms and governments, therefore, may not be coordinated in their deviation and production strategies. The coordination implied by our timing assumption is, in fact, not necessary for our results. We choose to focus on the case whereby a whole country deviates in the same period because we feel that this is the most interesting and most natural case. Hence, firms see only their own government's current-period tariff choice before they make their capacity decision.

Given the timing described above, the unconstrained capacity choice in each period can be written as a function of the other country's expected tariff in that period $Q_j(\tau^{ie}_{jt}) = (2\theta - \tau^{ie}_{jt})/3$. It is immediately apparent that the unconstrained capacity choice is declining in the expected tariff. From equation (5) output and capacity of good j in period t are given by:

$$Q_{jt} = \begin{cases} \text{Max}\{Q_j(\tau^{ie}_{jt}), Q_{j(t-1)}\} \text{ if } Q_{j(t-s)} > Q_{j(t-s-1)} \text{ for any } s \in \{1,\ldots,N\}, \\ Q_j(\tau^{ie}_{jt}) \qquad \text{otherwise.} \end{cases} \tag{8a}$$

We now turn to the depiction of the stationary tariff outcomes.

C. Free Trade.

From equations (6), (7) and (8a) we see that when free trade is expected: $Q^F_j = Q_j(0) = 2\theta/3$, the free-trade price is $P^j_j = P^i_j = (2A + B)/3$, and per-country consumption of each good is $D^j_j = D^i_j = \theta/3$. From equation (9) free-trade period-welfare is then $V^F = V^i(0,0,Q_i(0),Q_j(0)) = \theta^2/3 + \ell$.[11]

D. Markov-Nash Tariffs.

Each country's period-optimal tariff, τ^{im}, satisfies the following first-order-condition:

$$\frac{\partial V^i(\cdot,\tau^j,Q_i,Q_j)}{\partial \tau^i} = D^i_j(P^i_j(\tau^{im}))[1 - P^{i\prime}_j(\cdot)] + \tau^{im} \cdot D^{i\prime}_j(\cdot) = 0. \tag{10}$$

In the absence of export taxes and subsidies, $\tau^i = \tau^i_j$. We use this fact to simplify notation. From the expressions for P^i_j and D^i_j given in equations (6) and (7) we have $\frac{Q_j - \tau^{im}}{2} \cdot \frac{1}{2} - \tau^{im} \cdot \frac{1}{2} = 0$. It is straightforward to verify that $\partial^2 V^i / \partial \tau^{i^2} = -3/4$ so that equation (10) uniquely determines each country's best-response tariff: $\tau^{im} = Q_j/3$.

[11] Because exports are maximized and numeraire production is minimized in free trade, substituting Q^F into the total cost function that yielded (4) shows that positive numeraire production and a wage of one in both countries is ensured by $\ell > (2A^2 + 2AB - 4B^2)/9$.

There are four interesting things to notice about these best-response tariffs. First, they are increasing in the capacity choice of the trading partner. This indicates that the more a country is dependent on trade and specialized towards an export market, the higher the optimal tariff that can be levied against that country. Equivalently, greater import penetration generates a more protectionist optimal tariff. Second, the best response tariffs are stationary if and only if the capacity choices are stationary. Third, the best-response tariffs are not a function of the other country's tariff, therefore, they uniquely define the Markov-Nash-equilibrium tariffs. Fourth, from equations (6) and (7) it is evident that tariffs are never high enough to choke off all trade. Put another way, they are always below the prohibitive tariff level.[12]

E. Markov-Perfect Equilibrium.

As in a framework with fully reversible capacity, one subgame-perfect equilibrium for this dynamic tariff game is an infinite repetition of the static Nash equilibrium. In this benchmark case, firms and governments expect a Markov-Nash tariff in every period. These Markov-Nash tariffs are not conditioned on past tariff outcomes. A subgame-perfect equilibrium in these Markovian strategies is a **Markov-perfect equilibrium** (MPE.) If capacity is costlessly and immediately reversible, or if capacity is never added, then the physical environment, as described by the state variable, would look the same to the firms and the governments in every period. The unique MPE in this case would be the infinite repetition of the static Nash equilibrium. The irreversible capacity indicates that histories with greater installed capacity may generate different MPE outcomes. We now characterize this MPE set.

Lemma 1: *(i.) The unique MPE after any history with no capacity additions in the previous N periods or for the entire game is:* $Q^m_{jt} = Q^m_{j0} = Q_j(\tau^{im}_t(Q_j)) = 3\theta/5$ *and* $\tau^{im}_t = \tau^{im}_0 = \tau^{im}_t(Q_j(\tau^{im}_t)) = \theta/5$ *for all t.*

---

[12] The monotonic effect of Q_j on τ^{im} (point 1) results from our linear framework and facilitates derivation of our results. That the Markov-Nash equilibrium is in strictly dominant strategies (point 3) is a result of our linear partial-equilibrium framework and our omission of export policies. It is not necessary for our results, however, it does simplify their derivation. Point 4 indicates that autarky is not a one-shot equilibrium here. This difference with the case first noted by Dixit (1987, p.335) is as follows: by not considering export taxes, we ensure that prohibitive tariffs cannot be an equilibrium in weakly dominated strategies.

(ii.) After any other history (when capacity was last installed in period s) the unique MPE is:

$Q^{m}_{jt} = Q_{js} \geq Q^{m}_{j0} \geq Q_j(\tau^{im}_t(Q_{jt}))$ *and* $\tau^{im}_t(Q_{jt}) \geq \tau^{im}_0$ *for t* $\in$ *{s+1,…, s+N}.*

The proof of Lemma 1 is contained in Chisik (2001) and is available upon request from the author.

As in the standard, reversible capacity case, free trade Pareto dominates the **no trade agreement** MPE (the unique MPE for the entire game). To see this substitute $\{Q^{m}_{j0}, \tau^{im}_0\}$ into equation (6) and then equation (9) to yield period utility of $V^{im}_0 = V^i(\tau^{im}_0, \tau^{jm}_0, Q_i(\tau^{jm}_0), Q_j(\tau^{im}_0)) = 8\theta^2/25 + \ell < \theta^2/3 + \ell = V^F$.

The inefficiency here is a result of the asymmetric effect of a tariff on each country's welfare. Although a small tariff increases the levying country's welfare, it decreases the trading partner's welfare by a greater amount. To see this, note that world welfare ($=\Sigma_i V^i(\tau,\tau,Q,Q)$) is decreasing in a common tariff, τ. This tariff setting interaction is, therefore, reminiscent of a prisoners' dilemma, whereby each country benefits from a mechanism that enforces a more cooperative outcome. We now turn to such mechanisms.

F. Trade Agreement Strategies

A common form of **enforcement mechanism** encountered in the literature relies upon "grim" strategies. These history dependent strategies mandate that any deviation from the specified cooperative action will generate an infinite reversion to a punishment stage. A credible punishment threat relies upon continuation payoffs from a perfect equilibrium (often the most undesirable static Nash-equilibrium.) Adapted to a tariff setting framework these strategies specify a **most-cooperative** (or lowest) **self-enforcing tariff** (τ^{ic}) as a function of the discount rate and also a punishment tariff to be levied for the infinite future if a country deviates from the cooperative tariff. The punishment tariffs are often the Nash-equilibrium tariffs and the punishment stage can be considered as a trade-war.

The evolution of the state variable, here, naturally precludes reversion to a static Nash-equilibrium. A natural counterpart in stochastic games (see Ausubel and Deneckere (1987), Cave (1987) and Radner and Benhabib (1992)) is an infinite reversion to the worst MPE for the offending party. By Lemma 1, we know that there is a unique MPE following any deviation and we can utilize this MPE in describing a subgame-perfect enforcement mechanism. Let us call these modified grim strategies as the **trade agreement strategies**. To see explicitly the conditioning of these strategies upon history, first define $\Gamma^{i}_{t+1}(H^{t+1}) \to \tau^{i}_{t+1}$ as government i's action in period t+1. Similarly define country i's aggregate industry action as $\xi_{i(t+1)}(H^{t+1}, \tau^{i}_{t+1}) \to Q_{i(t+1)}$. Finally define a sequence of cooperative tariffs and capacities as $\tau^{ict} = \{ \tau^{ic}_{1}, \ldots, \tau^{ic}_{t} \}$ and $Q^{t}_{i}(\tau^{jct}) = \{ Q_{i}(\tau^{jc}_{1}), \ldots, Q_{i}(\tau^{jc}_{t}) \}$, respectively. The trade agreement strategies are defined as follows:

$$\Gamma^{i}_{t+1} = \tau^{ic}_{t+1} \text{ if } H^{t+1} = \{\tau^{ict}, \tau^{jct}, Q^{t}_{i}(\tau^{jct}), Q^{t}_{j}(\tau^{ict})\}$$

$$\tau^{im}_{t+1} \text{ otherwise;}$$

$$\xi_{i(t+1)} = Q_{i(t+1)}(\tau^{jc}_{t+1}) \text{ if } H^{t+1} = \{\tau^{ict}, \tau^{jct}, Q^{t}_{i}(\tau^{jct}), Q^{t}_{j}(\tau^{ict})\} \text{ and } \Gamma^{i}_{t+1} = \tau^{ic}_{t+1}$$

$$Q^{m}_{i(t+1)} \text{ otherwise.} \tag{12}$$

A potential difficulty here is the actions of the firms in the deviating country. The trade agreement strategies indicate that they revert immediately to trade war actions rather than waiting until the following period. Foreseeing the future retaliatory tariff that they will face, it is reasonable to expect that they would choose not to add capacity in the deviation period even when the cooperative tariff is declining in this period. If δ, or N, is small or the tariff cut is large, however, the discounted future losses from the excess capacity is less than the current gain and they could desire a small capacity expansion in the deviation period.[13] In any case, this desired increase, if it occurs at all, is generally very small, but allowing for its existence causes the resulting irreversible capacity to be a function of N and δ. This more complicated capacity expression complicates the analysis without changing the main results. We therefore assume the existence of an institutional constraint that precludes firms from adding capacity in

[13] It is shown in Chisik (2001) that the desired capacity, in the deviating country, in a deviation period is decreasing in N or δ and increasing in the size of the tariff reduction. Furthermore, if N or δ is large, or if the tariff cut is small, or if the existing capacity is well above Q^{m}_{0}, then the desired capacity does not increase at all. It is also shown that allowing for these possible capacity expansions does not change the main results. They can, however, slightly increase the pace of tariff liberalization.

the deviation period. Formally, this constraint may be considered as a threat by their government to levy an export tax on any output produced by capacity added in the deviation period. Alternatively, the remuneration package of the firm's decision makers may penalize excessive unused capacity or large profit margin reductions.

From the trade agreement strategies and from equation (8a), a **deviating** country's capacity in the deviation period (s) and in the first N-1 periods of a trade war is, therefore given by:

$$Q_{it}^{m} = Q_{i(s-1)}^{c} = Q_i\,(\tau_{(s-1)}^{jc}) = [2\theta - \tau_{(s-1)}^{jc}]/3 \ \text{ for } t \in \{s,\ldots, s+N-1\}. \tag{8b}$$

Capacity in the cooperating country is given by:

$$Q_{jt}^{m} = Q_{js}^{c} = Q_j\,(\tau_{s}^{ic}) = [2\theta - \tau_{s}^{ic}]/3 \ \text{ for } t \in \{s,\ldots, s+N\}. \tag{8c}$$

There are two forms of capacity asymmetries during a trade war. The first, and most important, is that the deviating country has less installed capacity than the cooperating country. The second is that the deviating country planned for the trade war one period earlier and, therefore, can remove their excess capacity one period earlier. Although a reasonable by-product of our model, this second asymmetry is in no way necessary for our results. Furthermore, as the extent of irreversibility (N) becomes large, this second asymmetry becomes vanishingly small.

IV. Non-Stationary Tariffs

A. Irreversibility and Unenforceability of Free Trade as a Stationary Tariff Outcome.

A notable feature of these strategies is that, when the physical environment does not change over time, they can support zero tariffs, or free trade, when the discount rate exceeds a critical level. It is interesting to note that if capacity irreversibility is of sufficiently long duration, then free trade cannot be supported, here, *initially*, for any value of the discount factor. To see this, we consider the extreme case of **complete irreversibility** (i.e. when N approaches infinity) and compare free trade to the asymmetric trade war that follows a deviation from free trade in the initial period.

Suppose, without loss of generality, that while country y expects free trade in the initial period and installs capacity accordingly, country x deviates from the free-trade agreement. Country x installs only Q_{x0}^{m} and faces a trade-war tariff of τ_0^{ym}. Country x levies a tariff of $\tau^{xm}(Q_y(0)) = 2\theta / 9$. Using these values in equations (6), (7) and then (9) yields punishment-period utility for country x of $V^x(\tau^{xm}, \tau^{ym}, Q_x(\tau^{ym}), Q_y(0)) = 451\theta^2 / 1350 + \ell > \theta^2 / 3 + \ell = V^F$. The asymmetric trade war generates greater welfare for country x than does free trade.[14] The punishment, in this case, is certainly no punishment at all. Put another way, the potential for opportunism is too great and there is no credible deterrent. It is, therefore, seen that when starting from an initial period, without a trade agreement, free-trade cannot be obtained in a single period. We state this above argument concisely as proposition 1:

Proposition 1: *If countries follow trade agreement strategies, and if there is complete irreversibility, then for all discount factors* $\delta \leq 1$, $\tau^F = 0$ *is not a stationary tariff outcome.*

When N is finite, free trade may be initially self-enforcing if the discount factor is sufficiently high. As shown in the next section, however, the critical discount factor that permits this outcome is an increasing function of N. For an intuitive explanation, note that a deviating country receives the higher asymmetric trade war payoff for the first N periods of a trade war and receives the lower no trade agreement payoff after that. As N increases, the undesirable part of the trade war occurs further in the future and a higher discount factor is required to deter this opportunistic behavior.

It is reasonable to expect that governments (or, at least, elected policy makers) have a relatively short expected time horizon and, therefore, place a low value on future payoffs so that free trade is

[14] That one country may prefer a trade war to free trade was also shown by Johnson (1953), Mayer (1981), Kennan and Riezman (1988), and McLaren (1997), among others. In each of these previous analyses exogenous asymmetries in country size yield this result rather than solely the endogenous capacity asymmetries as in the present case. Furthermore, this result is not dependant on our assumed functional forms. For example, in a Ricardian model with two goods, country x's Markov-tariff levied on imports from a specialized country y is extremely high. For reasonable elasticities of substitution and magnitudes of comparative advantage, country x would prefer an opportunistic asymmetric trade war over free trade in this case as well.

initially not self-enforcing even when irreversibility is of short duration.[15] In these cases, free trade, if obtained at all, must come about gradually. We now turn our attention to gradual liberalization outcomes.

B. Gradualism with Complete Irreversibility.

When free trade is not initially enforceable, tariffs lower than the no trade agreement MPE are self-enforcing. What is new, here, is that these lower tariffs result in higher installed capacity levels than Q^m_{j0}. The possible trade war outcome then differs in the following period, which causes the lowest self-enforcing tariff to be non-stationary.[16] To characterize this **cooperative tariff path,** and show that it exhibits gradualism, we first introduce the costs and benefits of adhering to the trade agreement. To develop intuition we restrict attention to the case of complete irreversibility in this section. We consider the more general case and provide a fuller analysis of these costs and benefits in section V.

Along the symmetric cooperative path $\tau^{ie}_t = \tau^{ic}_t = \tau^{jc}_t = \tau^c_t$, therefore, when abiding by the trade agreement in period s, the welfare in country i is:

$$\sum_{t=s}^{\infty}\delta^{t-s}V^{ic}_t = \sum_{t=s}^{\infty}\delta^{t-s}V^i_t(\tau^c_t) = \sum_{t=s}^{\infty}\delta^{t-s}V^i(\tau^{ic}_t,\tau^{jc}_t, Q_i(\tau^{jc}_t),Q_j(\tau^{ic}_t))\,. \tag{13}$$

Without loss of generality, suppose that country x is considering a deviation from the trade agreement in some period s. Clearly, the deviating government's optimal action is given by the Markov-Nash tariff: $\tau^{xm}_s = Q_y(\tau^c_s)/3$. They receive the deviation payoff for 1 period:

$$V^{xd}_s = V^{xd}(\tau^c_s, Q^c_{x(s-1)}) = V^{xd}(\tau^{xm}_s(Q_y(\tau^c_s)),\tau^c_s, Q^c_{x(s-1)}, Q_y(\tau^c_s)) = V^x(\tau^{xm}_s,\tau^c_s, Q^m_{xs}, Q_y(\tau^c_s)) \tag{14}$$

---

[15] An alternative justification for assuming that governments have a low discount factor in trade negotiations is as follows. The infinitely repeated game with discount factor, δ, can be considered as a finitely repeated game with a constant, and common knowledge, hazard rate that the game continues. In this interpretation $\delta = he^{-\rho L}$, where h is the hazard rate, ρ is the rate of time preference and L is the period length. Staiger (1995, pp. 1520-1521) explains that the period length can be thought of as the time required for observing and responding to the trading partner's policies. He provides compelling logic and some historical evidence that this time requirement is often lengthy, so that the government's discount factor is relatively low.

[16] When N = 0, there are no irreversibilities, the physical environment does not change over time, tariffs are not further reduced in later periods, and our model is similar to other repeated games encountered in the literature.

We now consider the trade war payoffs. For concreteness, again consider the payoff to country x when their period s deviation precipitated the trade war. When there is complete irreversibility ($N \to \infty$), these payoff can be written as:

$$\sum_{t=s+1}^{\infty} \delta^{t-s} V_s^{xm} = \sum_{t=s+1}^{\infty} \delta^{t-s} V^{xm}(\tau_s^c, Q_{x(s-1)}^c) = \\ \sum_{t=s+1}^{\infty} \delta^{t-s} V^x(\tau^{xm}(Q_{ys}^c(\tau_s^c)), \tau^{ym}(Q_{x(s-1)}^c), Q_{x(s-1)}^c, Q_{ys}^c(\tau_s^c)) = \sum_{t=s+1}^{\infty} \delta^{t-s} V^x(\tau_t^{xm}, \tau_t^{ym}, Q_{xt}^m, Q_{yt}^m) \tag{15}$$

For country X, the gain from deviating from the trade agreement in period s is:

$$\Psi_s^x = \Psi_s^x(Q_{x(s-1)}^c, \tau_s^c) = V_s^{xd} - V_s^{xc} \tag{16}$$

This gain must be balanced against the cost of a future trade war:

$$\Omega_s^x = \Omega_s^x(Q_{x(s-1)}^c, \tau_s^c, \delta) = \sum_{t=s+1}^{\infty} \delta^{t-s}[V_t^{xc} - V_s^{xm}] \tag{17}$$

Given the symmetry of the countries, the trade agreement is a sequence of cooperative tariffs $\{\tau_t^c\}$ that

$$\text{Maximize} \qquad \sum_{i \in \{X,Y\}} \sum_{t=s}^{\infty} \delta^{t-s} V_t^i(\tau_t^c) \tag{18}$$

subject to the constraint that the chosen cooperative tariffs and resulting capacity choices do not cause the gain from deviating from the agreement to be greater than the cost of a future trade war.

$$\Psi_s^i(Q_{i(s-1)}^c, \tau_s^c) \le \Omega_s^i(Q_{i(s-1)}^c, \tau_s^c, \delta), \quad i \in \{x,y\}, \forall s. \tag{19}$$

It is straightforward to verify that world welfare (as defined by equation 18), is a strictly decreasing concave function of the tariff rate and, therefore, is maximized by free trade. The trade agreement, therefore, specifies the lowest tariffs that satisfy the incentive constraint given by equation (19).

Having written the incentive constraint in terms of $Q_{i(s-1)}^c$ and τ_s^c we graph it in Figures 2 and 3. Note that Ψ_s^i is a strictly convex function of τ_s^c. Figure 2 considers the first period of the agreement, when irreversible investment is at the no-trade-agreement level. We see there that when $\tau_1^c = \tau_1^m$, there is no gain from deviating and $\Psi_1^i = 0$. Finally, we see that Ψ_1^i is negatively sloped at this no trade agreement tariff. These points are analytically established in the next section. We also show there that

Ω_s^i intersects Ψ_s^i twice and that $\Omega_s^i > \Psi_s^i$ between these two intersections. All tariffs in this interval are self-enforcing. The trade agreement coordinates the countries on the Pareto preferred, or lowest, self-enforcing tariff. This tariff is denoted as $\underline{\tau}_s^c$. One intersection is naturally at $\tau^m(Q_0^m)$. If $\delta > 1/3$, then the second intersection is at a lower tariff rate. Figure 2 also illustrates the result of Proposition 1.

FIGURES 2 AND 3 GO APPROXIMATELY HERE.

It is important to note that, because some tariff reduction is possible, Q_0^m is not an absorbing state of this dynamic game. In Figure 3 we see that if $Q_1^i > Q_0^{im}$, then Ω_2^i shifts up and Ψ_2^i shifts down. This slackening of the incentive constraint allows the lowest self-enforcing tariff to drop further in period 2 of the trade agreement. Hence, $\underline{\tau}_2^c < \underline{\tau}_1^c < \tau_0^m$. This tariff drop generates a further increase in irreversible capacity so that, by the same mechanism, tariffs are further reduced in each subsequent period. The above analysis is stated concisely in the following claim. We defer the formal development of a more general proposition until the next section.

CLAIM 1: *If governments value the future enough* ($\delta > 1/3$)*, then when starting at the no-trade-agreement MPE, there exists a self-enforcing cooperative tariff in the first period of a trade agreement that is strictly less than* $\tau^m(Q_0^m)$*. In the case of complete irreversibility, this initial tariff is strictly positive. This cooperative tariff generates an increase in the installed capacity that is verifiable at the end of the first period that, in turn, generates a further drop in tariffs in the second period. Tariffs continue to drop over time and eventually reach free trade iff* $\delta > 3/5$*.*

V. Gradualism in a General Case and Comparative Statics

A. Cooperation, Deviation, and Trade War Payoffs.

First consider the value to abiding by the agreement in period s. Using the common cooperative tariff τ_t^c and substituting equation (8a) into equation (7) yields $D_{jt}^{ic} = (\theta - 2\tau_t^c)/3, \quad D_{jt}^{jc} = (\theta + \tau_t^c)/3.$

Using equations (8a) and the above demand functions the period utility, equation (9), can be rewritten as:

$$V_t^{ic} = (Q_{it}^c)^2/2 + (D_{it}^{ic})^2/2 + (D_{jt}^{ic})^2/2 + \tau_t^c \cdot D_{jt}^{ic} + \ell = [\theta^2 - (\tau_t^c)^2]/3 + \ell. \quad (20)$$

Note that, along the cooperative path welfare is maximized at free-trade and $V^F = \theta^2/3 + \ell$.

Second, consider the payoff to a deviating country X. The demand for good y in each country satisfies: $D^{xm}_{ys} = Q^c_{ys}/3$; $D^{ym}_{ys} = 2Q^c_{ys}/3\tau^c_d$. Hence, consumer surplus and tariff revenue in the import sector in the deviating country x is $(D^{xm}_{ys})^2/2 + \tau^{xm}_s \cdot D^{xm}_{ys} = 3(Q^c_{ys})^2/18$. Export sector consumer surplus in the deviating country x is $D^x_{xs}(A - P^{xc}_{xs})/2 = (Q^c_{x(s-1)} + \tau^c_s)^2$. Although capacity is not increased, export sector profits are augmented by the improved terms of trade. From figure 4, we see that deviation period profits in country x are $Q^c_{x(s-1)}[P^{xc}_{x(s-1)} - B]/2 + [P^{xc}_{xs} - P^{xc}_{x(s-1)}]Q^c_{x(s-1)} = [(Q^c_{x(s-1)})^2 + Q^c_{x(s-1)}(\tau^c_{(s-1)} - \tau^c_s)]/2$. Using equations (8b) and (8c) to eliminate Q^c_{ys} and τ^c_{s-1} from these expressions for consumer surplus, tariff revenue, and profits, we can write the deviation period payoff solely as a function of $Q^c_{x(s-1)}$ and τ^c_s:

$$V^{xd}_s = [-189(Q^c_{x(s-1)})^2 + 216\cdot\theta\cdot Q^c_{x(s-1)} - 54\cdot\tau^c_s\cdot Q^c_{x(s-1)} + 31(\tau^c_s)^2 + 16\cdot\theta^2 - 16\cdot\theta\cdot\tau^c_s]/216 + \ell. \quad (21)$$

FIGURE 4 GOES APPROXIMATELY HERE.

Third, consider the future trade war payoffs to country x whose period s deviation generated the trade war. There are three components to these payoffs. In Chisik (2001) each component is derived in a manner similar to the derivation of equation (21). For the first N−1 periods there is an asymmetric trade war with capacities Q^c_{ys} and $Q^c_{x(s-1)}$ denoted as $V^{xm}_{s,N-1} = [-51(Q^c_{x(s-1)})^2 + 54\cdot\theta\cdot Q^c_{x(s-1)} + (2\theta - \tau^c_s)^2]/54 + \ell$.

In the Nth period of the trade war (period s+N), the deviating country x can reduce capacity to Q^m_0, while country y must maintain Q^c_{ys} until the following period: $V^{xm}_N = V^{xm}(\tau^c_s, Q^m_{x0}) = (\frac{351}{25}\theta^2 + [2\theta - \tau^c_s]^2)/54 + \ell$.

The trade war from period s+N+1 on is symmetric with capacities of Q^m_0. We then have: $V^{xm}_0 = V^{xm}(\tau^m_0, Q^m_{x0}) = \frac{8}{25}\theta^2 + \ell$. The period s value of the future trade war is equal to:

$$V^{xw}_s = \sum_{t=s+1}^{s+N-1}(\delta^{t-s}V^{xm}_{s,N-1}) + \delta^N V^{xm}_N + \sum_{t=s+N+1}^{\infty}(\delta^{t-s}V^{xm}_0) \quad (22)$$

B. A Gradual Approach to Free Trade: The Main Result

When capacity is reversible after N periods, the one period gain to deviating from the agreement (Ψ^i_s) does not change, however, the cost of a future trade war must be modified to:

$$\Omega^i_s = \Omega^i_s(Q^c_{i(s-1)}, \tau^c_s, \delta, N) = \sum_{t=s+1}^{\infty} \delta^{t-s} V^{ic}_t - V^{iw}_s \ . \tag{17a}$$

We can now state our key result regarding the shape of the incentive constraint.

Lemma 2: *(a.)* Ψ^i_s *is a strictly convex function of* τ^c_s *that attains a minimum at* $\tau^{c\,min}_s(Q^c_{i(s-1)}) > \tau^m_s(Q^F_i) \geq \tau^m_s(Q^c_{i(s-1)})$.

(b.) Ω^i_s *is a strictly increasing, and* Ψ^i_s *is a strictly decreasing, function of* $Q^c_{i(s-1)}$.

(c.) $\Omega^i_s(Q^m_0, \tau^c_s, \delta, N)$ *intersects* $\Psi^i_s(Q^m_0, \tau^c_s)$ *twice; and* $\Omega^i_s(Q^m_0, \tau^c_s, \delta, N)$ > $\Psi^i_s(Q^m_0, \tau^c_s)$ *between these intersections. One intersection occurs at* $\tau^m(Q^m_0)$ *and the other intersection is less than* $\tau^m(Q^m_0)$ *if **d** >* $\delta^c_0(N)$ ≤ *1/3, where* $\delta^c_0(N)$ *is increasing in N and lim*$_{N\to\infty}$ $\delta^c_0(N)$ *= 1/3.*

(d.) $\Psi^i_s(Q^m_0, 0) > \Omega^i_s(Q^m_0, 0, \delta, N), \forall\, \delta < \delta^F_0(N)$*, where* $\delta^F_0(N)$ *is increasing in N and lim*$_{N\to\infty}$ $\delta^F_0(N)$ *> 1.*

(e.) $\Omega^i_s(Q^F, 0, \delta, N) \geq \Psi^i_s(Q^F, 0)$ *iff* $\delta \geq \delta^F(N)$ ≥ *3/5, where* $\delta^F(N)$ *is decreasing in N and lim*$_{N\to\infty}$ $\delta^F(N)$ *= 3/5.*

The proof of Lemma 2 is contained in Chisik (2001), and is illustrated in Figures 2, 3, 5, 6.[17] Lemmas 2a and 2c are shown in Figure 5. By Lemma 2c, if $\delta >$ $\delta^c_0(N)$, then there are two intersections between Ω^i_s and Ψ^i_s. By Lemma 2a, the minimum of Ψ^i_s occurs at a tariff that is larger than the maximum chosen in any equilibrium, therefore, Ψ^i_s has a negative slope at the smaller intersection.

---

[17] The strict concavity of Ω^i (as illustrated, but never claimed) is not necessary for our results. What is necessary for our results is that Ψ^i cuts Ω^i from above at the smaller intersection. Given Lemmas 2d and 2e and the shape of Ψ^i this is true irrespective of the concavity of Ω^i.

Hence, when starting at the no-trade-agreement MPE, there exists a range of lower tariffs that are enforceable, if $\delta > \delta_0^c(N)$.

FIGURES 5 AND 6 GO APPROXIMATELY HERE

Lemma 2b is illustrated in Figure 3. It shows that an increase in the state variable, $Q_{i(s-1)}^c$, shifts the Ψ_s^i curve down and the Ω_s^i curve up. Lemma 2d provides a more constructive proof of Proposition 1 and is described graphically by Figure 2. Lemma 2e is represented by Figure 6. We see there that, when the state variable is equal to the free-trade capacity level, free trade is enforceable, if governments are sufficiently patient. From Lemma 2d and 2e and the shape of Ψ_s^i we see that Ψ_s^i intersects Ω_s^i from above at the lowest root of the quadratic equation $\Psi_s^i = \Omega_s^i$. Then from Lemma 2b, an increase in installed capacity in the previous period must reduce the lowest enforceable tariff in the current period. Finally, from equation (8b) we know that a previous period tariff reduction increases the previous period installed capacity. As a corollary of Lemma 2, with the above mentioned comparative static from equation (8b), we immediately obtain our main result.

Proposition 2: *If governments value the future enough* ($\delta > \delta_0^c(N)$), *then when starting at the no-trade-agreement MPE, there exists a self-enforcing cooperative tariff in the first period of a trade agreement that is strictly less than* $\tau^m(Q_0^m)$. *This cooperative tariff generates an increase in the installed capacity that is verifiable at the end of the first period that, in turn, generates a further drop in tariffs in the second period. Tariffs continue to drop over time and eventually reach free trade iff* $\delta > \delta^F(N)$.

C. Changes in the Extent of Irreversibility.

We now consider how the extent of irreversibility affects the obtainable trade agreement outcomes. A general theme throughout Lemmas 2c, 2d and 2e is that increasing irreversibility makes cooperation more difficult when capacity is near the no-trade-agreement level and makes it easier when capacity is near the free-trade level. In particular, Lemma 2c shows that $\delta_0^c(N)$ is increasing in N. The intuition for this comparative static result is as follows. A deviating country receives the higher asymmetric trade war payoff until capacity can be reversed and receives the lower symmetric no trade agreement payoff after that. An increase in the extent of irreversibility, therefore, decreases the cost of starting a trade war so that nations (and firms) act more cautiously. A corresponding increase in the common discount factor is then required to start the liberalization process. Lemma 2d shows that $\delta_0^F(N)$ is increasing in N. The intuition for this result is similar to that for Lemma 2c and is given in section IV.A above. Finally, Lemma 2e shows that $\delta^F(N)$ is decreasing in N. When capacity is maximized in both countries (and symmetric), an increase in the extent of irreversibility increases the cost of a future trade war. Cooperation is, therefore, easier to sustain and the critical value of the discount rate is reduced. The above argument is concisely summarized as:

Proposition 3: *An increase in the extent of irreversibility makes it harder to start a new trade agreement, but makes it easier to sustain a well-established agreement.*

D. Gradualism and Changes in the Gains From Trade.

A cost reducing technological improvement in the exporting country, or a demand increasing change in consumer tastes, translates into an increase in our index of the gains from trade (θ). If capacity is costlessly and instantly reversible (as in the usual case encountered in the literature), then the effect of θ on $\underline{\tau}_s^c$ is ambiguous. By varying the extent of irreversibility, or the arrival time of the increase in the gains from trade, our non-stationary framework allows us to make some non-ambiguous predictions about the effect of θ on $\underline{\tau}_s^c$.

Proposition 4: *An increase in the present and future gains from trade increases $\underline{\tau}_s^c$, if N is large. An anticipated future increase in the gains from trade decreases the current period $\underline{\tau}_s^c$, if N is small and increases $\underline{\tau}_s^c$, if N is large.*

Proposition 4 (which is proved in Chisik, 2001) relies on the capacity asymmetries during the first part of a trade war. First, consider the deviation period. From equation (8) we see that capacity is increasing in θ and, consequently, by equation (10) the Markov-Nash tariff also increases in θ. Now, the effect of any import tariff (and certainly the Markov-Nash tariff) is to shift a disproportionate amount of the gains from trade to the levying country. Hence, Ψ_s^i is increasing in θ. In a trade war, each country's Markov-Nash tariff usurps some of the gains from trade on their import good. During the first part of a trade war, when there are asymmetries in the irreversible capacity, the country with less irreversible capacity expropriates more surplus than is expropriated from them. The deviating country's cost of a trade war is, therefore, decreasing in θ when capacities are asymmetric. When capacities are symmetric (as in the later part of a trade war), each country's trade war tariff appropriates a symmetric, and equivalent, extra measure of the gains from trade in their import market, while reducing the overall level of trade. The value of this lost trade is, naturally, increasing in θ and, therefore, so is the cost of a symmetric trade war. Hence, the extent of irreversibility determines the effect of θ on the cost of a trade war. When N is large, much of the trade war is asymmetric and Ω_s^i is decreasing in θ. As the extent of irreversibility is reduced, more of the trade war is symmetric and Ω_s^i is increasing in θ. Putting this all together, we see, from figure 7, that if N is large, then a present and future increase in θ must increase $\underline{\tau}_s^c$.

FIGURE 7 GOES APPROXIMATELY HERE

Similarly, an expected future increase in the gains from trade, which alters only Ω_s^i, and has no effect on Ψ_s^i, must raise (lower) the current period $\underline{\tau}_s^c$, if N is large (small.) Furthermore, if N is large and θ increases starting in some period t, then $\underline{\tau}_s^c$ unambiguously rises in that period and in all future periods as well. Hence, if the extent of irreversibility is high enough, then an increase in the present and future gains from trade will raise present and future tariffs and perversely retard the gradualism process.

E. The Discount Factor and the Tariff Path.

In this section we characterize the tariff path as a function of the discount factor. We do not obtain general results and consider only the special case when there is complete irreversibility and when free trade is eventually obtained. In particular, we look at tariffs in the two periods immediately preceding the first period of zero tariffs and analyze the speed of tariff reduction. Because $Q^c_{i(s-1)}$ is a function of τ^c_{s-1} we can use equation (8b) to write equations (16) and (17a) solely as a function of the previous, current and future period cooperative tariffs.

$$\Psi^i_s(\tau^c_{s-1}, \tau^c_s) = \Psi^i_s(Q^c_{i(s-1)}(\tau^c_{(s-1)}), \tau^c_s) \tag{23}$$

$$\Omega^i_s(\tau^c_{s-1}, \tau^c_s, \delta, N) = \Omega^i_s(Q^c_{i(s-1)}(\tau^c_{(s-1)}), \tau^c_s, \delta, N) \tag{24}$$

If $\delta > \delta^F(N)$, then from Proposition 2 there exists a first period, T, when tariffs are zero and tariffs remain at zero for all future periods. Taking the limit of the incentive constraint in period T as N approaches infinity we obtain the following equation for tariffs in period T−1: $\Psi^i_T(\tau^c_{T-1}, 0) = \lim_{N\to\infty} \Omega^i_T(\tau^c_{T-1}, 0, \delta, N)$. The smallest root of this quadratic expression is the lowest tariff in period T−1 that is consistent with free trade in period T. After some simplification, this tariff can be written as: $\underline{\tau}^c_{T-1} = \dfrac{18+28\Delta(1) - 2\sqrt{270 + 330\Delta(1) + 60\Delta(1)^2}}{63+68\Delta(1)}\theta$, where $\Delta(1) = \delta/(1-\delta)$. There are three interesting things to notice about this expression. First, it is increasing in $\Delta(1)$ and, therefore in δ. Hence, when governments are more patient, the incremental tariff reductions are larger. Second $\underline{\tau}^c_{T-1} \gtreqless 0$ as $\Delta(1) \gtreqless 3/2$. This is in accordance with our earlier finding that free-trade is attainable if and only if $\delta \geq \lim_{N\to\infty} \delta^F(N) = 3/5$. Third, as $\Delta(1)$ goes to infinity, $\underline{\tau}^c_{T-1}$ approaches a limiting value of $[14-(60)^{1/2}]\theta/34 < \theta/5 = \tau^m_0$. This again indicates that attaining free-trade takes at least two rounds of tariff reductions.

Using the above result for $\underline{\tau}^c_{T-1}$, we can rewrite the incentive constraint in period T−1. This incentive constraint implicitly defines $\underline{\tau}^c_{T-2}$ solely as a function of the discount rate. Proceeding recursively we generate the entire gradual tariff path. The incentive constraint in period T - 1 can be written as: $\Omega^{id}_{T-1}(\tau^c_{T-2}, \underline{\tau}^c_{T-1}[\Delta(1)]) - \lim_{N\to\infty} \Omega^{ic}_{T-1}(\tau^c_{T-2}, \underline{\tau}^c_{T-1}[\Delta(1)], \delta, N)$. After some tedious algebra, it is possible to show that $\underline{\tau}^c_{T-2}$ is a strictly increasing function of $\Delta(1)$ that is positive iff $\Delta(1) \geq 3/2$.

Furthermore, $\underline{\tau}^{c}_{T-2} \le \theta/5 = \tau^{m}_{0}$ iff $\Delta(1) \le 3.105$. This indicates that if the future is valued highly enough, then free trade occurs after two rounds of tariff reductions If $\Delta(1) < 3.105$, then free-trade requires at least three rounds of tariff reductions. Finally, $\tau^{c}_{diff}[\Delta(1)] = \underline{\tau}^{c}_{T-2}[\Delta(1)] - \underline{\tau}^{c}_{T-1}[\Delta(1)]$ is a strictly increasing function of $\Delta(1)$ that is positive iff $\Delta(1) > 3/2$.

FIGURE 8 GOES APPROXIMATELY HERE

The above discussion is stated concisely as Proposition 5 which is illustrated in Figure 8 and proved in Chisik (2001).[18] Figure 8 plots $\underline{\tau}^{c}_{T-2}[\Delta(1)]$ and $\underline{\tau}^{c}_{T-1}[\Delta(1)]$ for the case of $\theta = 1$.

Proposition 5: *For the case of complete irreversibility, if* $\Delta(1) \ge 3/2$*, then free trade eventually occurs and takes at least two rounds of tariff reductions. If* $\Delta(1) < 3.105$*, then it takes at least three rounds of tariff reductions. If the future is valued highly, then incremental tariff reductions are larger and free trade occurs more quickly. As the discount rate is reduced to* $\delta^{F}(N)$*, the number of tariff reductions required to reach free trade become infinite.*

VI. Conclusion

In this paper we show that gradualism may arise in international trade agreements in the presence of capacity irreversibility and suggest that this capacity irreversibility may also be considered as a proxy for many types of trade-partner specific capital. We first consider the case where capacity is fixed in perpetuity and find that free trade, if it occurs, must then be a gradual process. Even if capacity is eventually reversible, then when the future is not sufficiently valued, free trade cannot occur in an initial period. In either case, when free trade is not initially self-enforcing, there exists a self-enforcing tariff that is lower than the no trade agreement level. This initial tariff reduction leads to a small expansion of each country’s export sector. The irreversibility of this expansion changes the incentive constraint in the next period and reduces the lowest obtainable self-enforcing tariff in the next period.

We next analyze how changes in the extent of capacity irreversibility and in our index of the gains from trade effect the tariff path. We show that an increase in the extent of irreversibility makes it more difficult to start a new trade agreement but makes it easier to sustain a well-established one. We also establish that if the extent of irreversibility is sufficiently high, then an increase in the gains from trade can perversely retard the gradualism process. Finally, we characterize the tariff path for the case when free trade is eventually enforceable. As a dynamic counterpart to stationary folk theorem results, we show that the speed of liberalization, as given by the size of tariff reductions, is increasing in the discount factor.

A potential limitation of our analysis is related to the extent of sectors covered by the trade agreement. We suggested that the evolution of a single export sector and import tariff can serve as a proxy for many differing sectors and tariffs. Although we believe that this intuition translates to a richer model, there are still many unanswered questions. In particular, how do the degrees of comparative advantages and capacity irreversibilities affect the order of sector liberalization and the pace of tariff reductions? Although this topic represents further research we use it to motivate a discussion of the bicycle effect.

As trade relationships mature, they encompass more difficult sectors of the economy and trade negotiations are more likely to stall. In the GATT, and in many regional agreements, these gradually longer negotiation rounds always produce some form of last-minute agreement, if no other substantial concessions are obtained. These face-saving compromises are considered symptomatic of what Bhagwati (1988) termed the bicycle effect: If nations stop pedaling on the free-trade-bicycle, then they will fall off. That is, if trade negotiations end so that there are no further tariff reductions, then the bicycle effect predicts that the current degree of liberalization is no longer enforceable.

---

[18] If the per period capacity costs are considered solely as loan repayment, then capacity decisions are a function of the discount rate and the analysis in Proposition 5 is less clear. Hence, Proposition 5 is only technically valid when capacity costs stem from explicit and implicit contractual obligations. I would like to thank Rick Bond for bringing this point to my attention.

In our model, this corresponds to a cessation of all negotiations and, therefore, all further tariff reductions beginning in some period, s. With the expected tariff $\tau_t^e \geq \underline{\tau}_{s-1}^c$ for all future t, we then ask if $\underline{\tau}_{s-1}^c$ is enforceable in period s.

There are two opposing effects at work here. The first is the termination of the anticipated reduction in future cooperative tariffs. If these tariffs do not drop over time, then the cost of a trade war decreases. By itself, this effect increases $\underline{\tau}_s$ and it suggests that the current tariff cannot be maintained. The other effect is the increase in irreversible capacity in period s–1. This second effect mitigates the first effect so that the overall effect on $\underline{\tau}_s$ is ambiguous and $\underline{\tau}_s = \underline{\tau}_{s-1}^c$ is possible. When the irreversibility constraint is no longer binding the second effect is irrelevant and tariffs must increase. The extent of irreversibility determines how long the bicycle can be balanced while not pedaling forward. Once capacity is reversible it must topple over.[19]

This result suggests that the bicycle effect is more pronounced in newer trading relationships or when countries are less integrated with their trading partners. Although not modeled here, it is possible that more established trading relationships permit trade-partner-specific investments with a greater extent of irreversibility. These more established relationships can, therefore, more easily withstand the occasional disagreements and stalemates that must be avoided when nurturing younger, more fragile relationships.

---

[19] It is instructive to compare our discussion of the bicycle effect with the results contained in Staiger (1995) and Furusawa and Lai (1999). In all three papers the (tariff raising) loss of future liberalization gains must be contrasted with the (tariff lowering) change in the state variable. This tariff-raising effect dominates in Staiger (except in the final period of liberalization) and there is evidence of a bicycle effect. In Furusawa and Lai, in the event of a trade war, workers must pay an additional adjustment cost to re-enter the import-competing sector they just left, which increases the state variable (tariff-lowering) effect and they find no evidence of the bicycle effect. In our framework, as long as capacity is irreversible the dominating effect on tariffs is dependant on our parameterization. The key difference in our model is that irreversibility ends after N periods, so that the state variable effect vanishes and the bicycle effect is eventually evidenced.

References

Admati, A., and M. Perry, "Joint Projects without Commitment," Review of Economic Studies 1991, 58, pp. 259-276.

Ausubel, L., and R. Deneckere, "One is Almost Enough for Monopoly," RAND Journal of Economics, Summer 1987, 18, pp. 255-274.

Benhabib, J., and R. Radner, "The Joint Exploitation of a Productive Asset: A Game-theoretic Approach," Economic Theory 1992, 2, pp.155-190.

Bagwell, K., "Commitment and Observability in Games," Games and Economic Behavior, 1995, 8, pp. 271-280.

Bagwell, K., and R. Staiger, "A Theory of Managed Trade," American Economic Review, September 1990, 80:4, pp. 779-795.

Bagwell, K., and R. Staiger, "Multinational Tariff Cooperation During the Formation of Regional Free Trade Areas," International Economic Review, May 1997, 38:2, pp. 291-319.

Bhagwati, J., *Protectionism,* Cambridge, MA: MIT Press; 1988.

Bond, E., and J.-H. Park, "Gradualism in Trade Agreements with Asymmetric Countries," Review of Economic Studies, forthcoming.

Cave, J., "Long-term Competition in a Dynamic Game: The Cold Fish-War," Rand Journal of Economics Winter, 1987, 18, pp. 596-610.

Chisik, R., "Gradualism in Free Trade Agreements: A Theoretical Justification," 2001 manuscript.

Devereux, M, "Growth, Specialization, and Trade Liberalization," International Economic Review, August 1997, 38:3, pp. 565-585.

Dixit, A., "Strategic Aspects of Trade Policy," in Bewley (ed.) *Advances in Economic Theory: 5th World Congress*, New York: Cambridge University Press; 1987, pp. 329-362.

Falvey, R. and C. Kim, "Timing and Sequencing Issues in Trade Liberalization," Economic Journal, July 1992, 102:413, pp. 908-924.

Furusawa, T., and E. Lai, "Adjustment Costs and Gradual Trade Liberalization," Journal of International Economics, December 1999, 49:2, pp. 333-361.

Helpman E., and P. Krugman, *Trade Policy and Market Structure*, Cambridge, MA: MIT Press; 1989.

Hungerford, T., "GATT: A Cooperative Equilibrium in a Noncooperative Trading Regime?" Journal of International Economics, November 1991, 31:3-4, pp. 357-369.

Irwin, D., "Changes in U.S. Tariffs: The Role of Import Prices and Commercial Policies," American Economic Review, September 1998, 88:4, pp. 1015-1026.

Lapan, H., “The Optimal Tariff, Production Lags, and Time Consistency,” American Economic Review, June 1988, 78:3, pp. 395-401.

Johnson, H., “Optimum Tariffs and Retaliation,” Review of Economic Studies, 1953-54, 21:2, pp.142-153.

Kennan, J. and R. Riezman, “Do Big Countries Win Tariff Wars?” International Economic Review, February 1988, 29:1, pp. 81-85.

Lapham, B., “Tariffs and Welfare in a Dynamic Differentiated Oligopoly,” Review of International Economics, 1995, 3:1, pp. 60-74.

Maggi, G., “The Value of Commitment with Imperfect Observability and Private Information,” RAND Journal of Economics, Winter 1999, 30:4, pp.555-574.

Mayer, W., “Theoretical Considerations on Negotiated Tariff Adjustments,” Oxford Economics Papers, March 1981, 33:1, pp.135-153.

McLaren, J., “Size, Sunk Costs, and Judge Bowker’s Objection to Free Trade,” American Economic Review, June 1997, 87:3, pp. 400-420.

Mehlum, H., “Why Gradualism,” The Journal of International Trade and Economic Development, 1998, 7:3, pp. 279-297.

Mussa, M. , “The Adjustment Process and the Timing of Trade Liberalization,” in A. Choksi and D. Papageorgiou (eds.) *Economic Liberalization in Developing Countries,* Oxford: Basil Blackwell, 1986.

Riezman, R., “Dynamic Tariffs with Asymmetric Information,” Journal of International Economics, May 1991, 30:3-4, pp. 267-283.

Rosen, H., “The US-Israel Free Trade Area Agreement: How Well Is It Working and What Have We Learned?” in Schott (ed.) *Free Trade Areas and U.S. Trade Policy*, Washington D.C.: Institute for International Economics; 1989.

Rowat, M., Lubrano, M., and R. Porrata, Jr., “Competition Policy and MERCOSUR,” World Bank technical paper 385, 1997.

Staiger, R., “A Theory of Gradual Trade Liberalization,” in J. Levinsohn, A. Deardorff, and R. Stern (eds.) *New Directions in Trade Theory*, Ann Arbor, MI: University of Michigan Press; 1995, pp.249-284.

Staiger, R., “International Rules and Institutions for Trade Policy.” in G. Grossman and K. Rogoff (eds.) *HandBook of International Economics, Vol. 3*, Amsterdam: North-Holland; 1995, pp.1495-1551.

Staiger, R. and G. Tabellini, “Discretionary Trade Policy and Excessive Protection,” American Economic Review, December 1987, 77:5, pp. 823-837.

Appendix A: Proofs. (Available by request.)

Proof of Lemma 1: (i.) In any period in which capacity was not expanded in the previous N periods, as well as in the initial period, the irreversibility constraint is not binding, so that $Q_{jt} = Q_j(\tau_t^{im})$. The physical environment is identical and, therefore, the possible MPE outcomes are identical in each such period. Substituting $\tau_0^{im} = Q_{j0}/3$ into equation (8a) yields $Q_{j0}^m = Q_j(\tau_0^{im}) = 3\theta/5$. This capacity implies a tariff of $\tau_0^{im} = \tau_0^{im}(Q_{j0}(\tau_0^{im})) = \theta/5$. Expecting the same tariff in each subsequent period, firms neither desire to add or reduce capacity and, therefore, the Markov-Nash tariff is the same in each subsequent period. Hence, these initial tariff and capacity pairs constitute a MPE in each subsequent period as well.

(ii.) After any other history, with $Q_{js} \geq Q_{j0}^m$, it must be the case that $\tau_{js}^{ie} \leq \tau_0^{im}$. Expecting a Markov-Nash tariff in, say, period s+1, the desired capacity becomes $Q_j(\tau_{(s+1)}^{im}(Q_{js})) \leq Q_{js}$. Firms, therefore, do not add capacity and $Q_{j(s+1)}^m = Q_{js}$, which implies a tariff of $\tau_{(s+1)}^{im} = Q_{js}/3 \geq Q_{j0}^m/3$. Using the same argument as in the proof of part (i), these tariff and capacity pairs constitute a MPE in each subsequent period $t \in \{s+2,\ldots,s+N\}$.

Derivation of Trade War Payoffs for the Deviating Country

In the trade war generated by country x's deviation, country x continues to levy a Markov-Nash tariff, therefore, their import sector consumer surplus and tariff revenue during the trade war is the same as in the deviation period. Country x's exports now face a Markov-Nash tariff and their export sector consumer surplus is equal to: $(D_{xs}^{xm})^2/2 = 4(Q_{x(s-1)}^c)^2/18$. Producer surplus (for the first N–1 periods) is slightly more complicated, because the desired industry capacity for competitive profit-maximizing firms facing the trade-war prices is less than the actual industry capacity. In Figure 4 we denote $\hat{Q}_{xt}^m$ as the desired capacity for the given price ($P_{xt}^{xm} = B + \hat{Q}_{xt}^m$) Producer surplus in a trade war is then given by: $\hat{Q}_{xt}^m \cdot [P_{xt}^{xm} - B]/2 - [P_{x(s-1)}^{xc} - P_{xt}^{xm}] \cdot [Q_{x(s-1)}^c - \hat{Q}_{xt}^m]/2$. Using equations (6) and (10) it is easy to see that: $\hat{Q}_{xt}^m = [3\theta - 2Q_{x(s-1)}^c]/3$. Furthermore, using equations (6) and (10) and then (8b) yields: $P_{x(s-1)}^{xc} - P_{xt}^{xm} = [\tau_t^{ym} - \tau_{(s-1)}^c]/2 = [Q_{x(s-1)}^c - 3\tau_{(s-1)}^c]/6 = [5Q_{x(s-1)}^c - 3\theta]/3$. Substituting these equations into equation (9) and then using equation (8c) to make a substitution for Q_{ys}^c yields period utility for the first N–1 periods of the trade war as: $V_{s,N-1}^{xm} = [-51(Q_{x(s-1)}^c)^2 + 54 \cdot \theta \cdot Q_{x(s-1)}^c + (2\theta - \tau_s^c)^2]/54 + \ell$ for $t \in \{s+1,\ldots, s+N-1\}$. In period N of the trade war capacity is not constrained in the deviating country and the above equation can be written with Q_0^m in place of $Q_{x(s-1)}^c$ yielding:

$V_N^{xm} = V^{xm}(\tau_s^c, Q_{x0}^m) = (\frac{351}{25}\theta^2 + [2\theta - \tau_s^c]^2)/54 + \ell$, for t = s+N. After period N of the trade war capacity is unconstrained in both countries and the above equation can be written with $\tau_0^m = \theta/5$ in place of τ_s^c. We then have: $V_0^{xm} = V^{xm}(\tau_0^m, Q_{x0}^m) = \frac{8}{25}\theta^2 + \ell$ for t > s+N. The trade war payoffs can then be written as:

$$V_s^{xw} = \sum_{t=s+1}^{S+N-1}(\delta^{t-s}V_{s,N-1}^{xm}) + \delta^N V_N^{xm} + \sum_{t=s+N+1}^{\infty}(\delta^{t-s}V_0^{xm}) =$$

$$[\Delta(1) - \Delta(N)]V_{s,N-1}^{xm} + [\Delta(N) - \Delta(N+1)]V_N^{xm} + [\Delta(N+1)]V_0^{xm},$$

where $\Delta(t) \equiv \delta^t/(1-\delta)$ is introduced to simplify notation. We show in appendix B that $\Delta(t)$ is decreasing in t, increasing in δ and , for $t' > t$, $\Delta(t, t') = \Delta(t) - \Delta(t')$ is increasing in δ.

Proof of Lemma 2: First, using equations (20) and (21) we can rewrite equation (16) as:

$$\Psi_s^x = [-189(Q_{x(s-1)}^c)^2 + 216\cdot\theta\cdot Q_{x(s-1)}^c - 54\cdot\tau_s^c\cdot Q_{x(s-1)}^c - 56\theta^2 - 16\cdot\theta\cdot\tau_s^c + 103(\tau_s^c)^2]/216\,.$$

Substituting the above components of V_s^{xw} into equation (22) and combining with equation (20) we can rewrite equation (17) as:

$$\Omega_s^x = \Big([350\theta^2 + 1275(Q_{x(s-1)}^c)^2 - 1350\cdot\theta\cdot Q_{x(s-1)}^c]\Delta(1,N) + [19\Delta(N+1) - \Delta(1)]\theta^2$$
$$+ [100\cdot\theta\cdot\tau_s^c - 25(\tau_s^c)^2]\Delta(1,N+1) - 450\sum_{t=s+1}^{\infty}\delta^{t-s}(\tau_t^c)^2 \Big)/1350.$$

Second, note that $2\theta/3 = Q^F \geq Q_{i(s-1)}^c \geq Q_{i0}^m = 3\theta/5$ and $\tau_s^c \leq \tau_s^m = Q_{i(s-1)}^c/3 \leq 2\theta/9$ in any equilibrium. We use these bounds on $Q_{i(s-1)}^c$ and τ_s^c repeatedly below.

(a) $\dfrac{\partial\Psi_s^i}{\partial\tau_s^c} = \frac{1}{216}[-54Q_{i(s-1)}^c - 16\theta + 206\tau_s^c] < 0\,.$

To see that Ω_s^{id} is convex in τ_s^c, note that $\dfrac{\partial^2\Psi_s^i}{\partial(\tau_s^c)^2} = \dfrac{206}{216} > 0\,.$

Argmin $\Psi_s^i(Q_{i(s-1)}^c,\cdot) = (54Q_{i(s-1)}^c + 16\theta)/206 \geq (54(3\theta/5) + 16\theta)/206 = 242\theta/1030 > 2\theta/9.$

(b) $\dfrac{\partial\Omega_s^i}{\partial Q_{i(s-1)}^c} = \dfrac{\Delta(1,N)}{54}[102Q_{i(s-1)}^c - 54\theta] > 0\,;\ \dfrac{\partial\Psi_s^i}{\partial Q_{i(s-1)}^c} = \dfrac{1}{216}[-378Q_{i(s-1)}^c + 216\theta - 54\tau_s^c] < 0\,.$[1]

[1] The intuition for this important result is as follows. An increase in τ^c has no effect on capacity in the deviating country, but decreases capacity in the surprised country. For this reason, Ω^i is initially increasing in τ^c. Similarly, Ψ^i is increasing in irreversible capacity in the surprised country and is, therefore, decreasing in τ^c. That V^w is decreasing, and, therefore, that Ω^i is increasing, in $Q_{(s-1)}$ is the main point of the paper and has been explained

(c) In the case of static expectations ($\tau_t = \tau_{t+1}, \forall t$),

$$\Omega_s^i(Q_0^m, \tau_s^c, \delta, N) - \Psi_s^i(Q_0^m, \tau_s^c) = \{[100\Delta(N+1) - 1900\Delta(1) - 2575](\tau_s^c)^2 +$$

$$[400\Delta(1, N+1) + 1210]\theta\tau_s^c + [76\Delta(N+1) - 4\Delta(1) - 139]\theta^2\}/5400. \qquad \text{(AP1)}$$

The above quadratic form is strictly concave in τ_s^c, therefore, if it has two roots, then it is positive between the roots. Solving for the roots yields

$$\tau_s^c = \left\{\frac{139 + 4\Delta(1) - 76\Delta(N+1)}{515 + 380\Delta(1) - 20\Delta(N+1)}\theta, \frac{103 + 76\Delta(1) - 4\Delta(N+1)}{515 + 380\Delta(1) - 20\Delta(N+1)}\theta\right\}.$$

The larger root is equal to $\tau^m(Q_0^m) = \theta/5$ and the smaller root is smaller than $\theta/5$ if:

$$2\delta^{N+1} + 3\delta - 1 > 0. \qquad \text{(AP2)}$$

Solving AP2 with equality implicitly defines a $\delta_0^c(N)$. Hence, because AP2 is strictly increasing in δ, it is satisfied for all $\delta > \delta_0^c(N)$. Because the implicit expression for $\delta_0^c(N)$ satisfies the conditions for the implicit function theorem, we have $\partial\delta_0^c(\cdot)/\partial N = -[2\delta^{N+1}\ln(\delta)]/[(N+1)2\delta^N + 3] > 0$. Furthermore, $\lim_{N\to\infty}[2\delta^{N+1} + 3\delta - 1] = 3\delta - 1$, so that $\lim_{N\to\infty}\delta_0^c(N) = 1/3$.

As it turns out, static expectations overestimates $\delta_0^c(N)$. If gradualism is expected ($\tau_t > \tau_{t+1}, \forall t$), then the first term in the above quadratic (AP1) is smaller in absolute value and the lower root takes on a smaller value, therefore, the lower bound of δ that permits tariff reductions is less than our derived $\delta_0^c(N)$. It is for this reason that Lemma 2c is only stated as a sufficient condition.

(d) $\Psi_s^i(Q_0^m, 0) = 139\theta^2/5400 > [19\Delta(N+1) - \Delta(1)]\theta^2/1350) = \Omega_s^i(Q_0^m, 0, \delta, N)$ iff $139 - 135\delta - 76\delta^{N+1} > 0$. (Note that this expression can also be derived by setting the smaller root of AP1 less than zero.) Solving this polynomial with equality implicitly defines a $\delta_0^F(N)$. Because this polynomial is decreasing in δ, it is, therefore, satisfied for all $\delta < \delta_0^F(N)$. Again, the implicit function theorem can be used to show that $\partial\delta_0^F(\cdot)/\partial N = -[-76\delta^{N+1}\ln(\delta)]/[-76(N+1)\delta^N - 135] > 0$. Furthermore,

elsewhere. What is less obvious is that V^d, and therefore Ψ^i, is decreasing in $Q_{(s-1)}$. This last derivative is small in magnitude, (and if it vanished it would have no effect on our results), however, the reason for its occurrence is interesting and can be described as follows. The competitive firms fail to internalize the negative external effect that their increased output has on the export market price and, therefore, produce more than the welfare maximizing quantity. (Hence, an export tax is the optimal policy for a competitive export industry.) This occurs even when exports are minimized as in the no-trade-agreement MPE. Although tariff reductions increase profits, they shift a greater percentage of output towards the export market. This second, indirect effect is more pronounced when output is larger so that Ψ^i is decreasing in $Q_{(s-1)}$.

$\lim_{N\to\infty}[139 - 135\delta - 76\delta^{N+1}] = 139 - 135\delta$, so that $\lim_{N\to\infty}\delta_0^F(N) > 1$. This last result can also be seen directly as $\Psi_s^i(Q_0^m, 0) = 139\theta^2/5400 > 0 > -\Delta(1)\theta^2/1350) = \lim_{N\to\infty}\Omega_s^i(Q_0^m, 0, \delta, N)$.

(e) $\Omega_s^i(Q^F, 0, \delta, N) = [50\,\Delta(1) - 53\Delta(N) + 57\,\Delta(N+1)]\theta^2/4050 \geq \theta^2/54 = \Psi_s^i(Q^F, 0)$ iff $125\delta - 53\delta^N + 57\delta^{N+1} - 75 \geq 0$. Now this polynomial is increasing in δ, so it is satisfied for all $\delta \geq \delta^F(N)$, where $\delta^F(N)$ satisfies this polynomial with equality. When N approaches ∞, $\delta^F(N)$ approaches 3/5 and when N= 0, $\delta^F(N) = 128/182$. The conditions for the implicit function theorem are again satisfied so that $\partial\delta^F(\cdot)/\partial N = -[(57\delta^{N+1} - 53\delta^N)\ln(\delta)]/[125 + 57(N+1)\delta^N - 53N\delta^{N+1}] < 0$ iff $\delta^F(N) < 53/57$, which is true by the previous sentence.

Proof of Proposition 4: The proof consists of noting the effect of θ on Ω_s^i and on Ψ_s^i.

$$\frac{\partial\Omega_s^i}{\partial\theta} = \{[700\theta - 1350\,Q_{i(s-1)}^c]\Delta(1, N) + [100\,\tau_s^c]\Delta(1, N+1) - 2\theta\Delta(N) + 38\theta\Delta(N+1)\}/1350.$$

$$\lim_{N\to\infty}\frac{\partial\Omega_s^i}{\partial\theta} = [700\theta - 1350\,Q_{i(s-1)}^c + 100\,\tau_s^c]\Delta(1)/1350 < 0.$$

When capacity is instantly reversible (N = 0), the cost of a trade war is

$\sum_{t=s+1}^{\infty}\delta^{t-s}(V_t^{ic} - V_0^{im}) = \sum_{t=s+1}^{\infty}\delta^{t-s}[25\theta^2 - (\tau_t^c)^2 - 24\theta^2]/75$, which is clearly increasing in θ. Furthermore, $\frac{\partial^2\Omega_s^i}{\partial\theta\partial N} = \{[1350\,Q_{i(s-1)}^c - 702\theta]\frac{\partial\Delta(N)}{\partial N} + [38\theta - 100\,\tau_s^c]\frac{\partial\Delta(N+1)}{\partial N}\}/1350 < 0$. This last result occurs, because $\partial\Delta(t)/\partial t = \Delta(t)\ln(\delta) \leq 0$ and $\delta \leq 1$. Hence, there exists a critical value of N, which is strictly positive, such that Ω_s^i is increasing (decreasing) in θ as N is below (above) this critical value.

Finally, we have $\frac{\partial\Psi_s^i}{\partial\theta} = [216\,Q_{i(s-1)}^c - 112\theta - 16\,\tau_s^c]/216 > 0$, so that Ψ_s^i is increasing in θ. Referring to Figure 7 immediately yields the statement of the Proposition.

Proof of Proposition 5: First, note that $\Psi_T^i(\tau_{T-1}^c, 0) - \lim_{N\to\infty}\Omega_T^i(\tau_{T-1}^c, 0, \delta, N) =$

$\{[12 - 2\Delta(1)]\theta^2 - [63 + 68\Delta(1)](\tau_{T-1}^c)^2 + [36 + 56\Delta(1)]\theta\tau_{T-1}^c\}/648 \leq 0$ Solving this quadratic yields:

$$\tau_{T-1}^c = \frac{18+28\Delta(1) - 2\sqrt{270 + 330\Delta(1) + 60\Delta(1)^2}}{63+68\Delta(1)}\theta.$$

$$\partial\underline{\tau}^{c}_{T-1}(\cdot)/\partial\Delta(1) = \frac{5(1593+1488\Delta(1)+[270+330\Delta(1)+60\Delta(1)^2]^{1/2})}{[63+68\Delta(1)^2][270+330\Delta(1)+60\Delta(1)^2]^{1/2}}\theta > 0\ .$$

Evaluating $\underline{\tau}^{c}_{T-1}(\Delta(1))$ at $\Delta(1) = 3/2$ yields $2[9 + 14(3/2) - (270 + 330(3/2) + 60(3/2)^2)^{1/2}]\theta/[63 + 68(3/2)] = 2[30 - 30]/165 = 0$. For the third claim regarding $\underline{\tau}^{c}_{T-1}$, note that $\lim_{\Delta(1)\to\infty}\underline{\tau}^{c}_{T-1} = (14-\sqrt{60})\theta/34$.

The proofs of the claims regarding τ^{c}_{T-2} are straightforward, however, they require manipulation of very lengthy derivatives and are omitted. They are available on request from the author.

Appendix B: Deviation Period Capacity Decisions. (Available by request.)

In this appendix, we analyze the deviation period capacity choices of firms when their options are not limited by institutional constraints. This capacity is denoted by $Q_s = \text{Max}\ \{Q^D, Q_{(s-1)}\}$, where Q^D is the desired capacity in the deviation period when there are no institutional or irreversibility constraints. We now investigate Q^D to discover when it is greater than $Q_{(s-1)}$. If firms add capacity in the deviation period, then the industry marginal revenue is $(2A - Q^D - \tau_s^c)/2$ in the deviation period and $\Delta(1, N)(2A - Q^D - Q^D/3)/2$ for the first N periods of the trade war, where $\Delta(1, N) = \Delta(1) - \Delta(N) = (\delta - \delta^N)/(1 - \delta)$ is introduced to economize on notation. The intertemporal industry marginal revenue from this capacity expansion is, therefore, equal to $(6A[1 + \Delta(1, N)] - 3\tau_s^c - Q^D[3 + 4\Delta(1, N)])/6$. The intertemporal industry marginal cost is $(1 + \Delta(1, N))(B + Q^D)$. Equating the intertemporal marginal revenue to the intertemporal marginal cost yields: $Q^D = [6\theta(1 + \Delta(1, N)) - 3\tau_s^c]/[9 + 10\Delta(1, N)]$.

It is straightforward to verify that:

(a.) $\partial Q^D/\partial \tau_s^c = -3/[9 + 10\Delta(1, N)] < 0$;

(b.) $\partial Q^D/\partial\Delta(1, N) = [-6\theta + 30\tau_s^c]/[9 + 10(\Delta(1, N)]^2 \leq 0$; (because $\tau_s^c \leq \theta/5$)

(c.) $\partial\Delta(N)/\partial N = \Delta(N)\ln(\delta) \leq 0$; (because $\delta \leq 1$)

(d.) $\partial\Delta(1, N)/\partial N = -\Delta(N)\ln(\delta) \geq 0$;

(e.) $\partial\Delta(N)/\partial\delta = [N\delta^{N-1} - (N-1)\delta^N]/(1-\delta)^2 \geq 0$;

(f.) $\partial\Delta(N+1)/\partial\delta - \partial\Delta(N)/\partial\delta = N\delta^{N-1}[2\delta - \delta^2 - 1]/(1-\delta)^2 \leq 0$; (because $2\delta - \delta^2 - 1 \leq 0$ for all $\delta \in [0, 1]$

(g.) $\partial\Delta(1, N)/\partial\delta \geq 0$; (because the result in part f holds for all N)

(h.) $\partial Q^D/\partial N) = [\partial Q^D/\partial\Delta(1, N)]\partial\Delta(1, N)/\partial N \leq 0$; (from parts b and d)

(i.) $\partial Q^D/\partial\delta) = [\partial Q^D/\partial\Delta(1, N)]\cdot\partial\Delta(1, N)/\partial\delta \leq 0$; (from parts b and g).

Now, firms do not desire to add capacity in the deviation period if $Q^D \leq Q_{(s-1)}$, which is equivalent to: $(\delta - \delta^N)/(1 - \delta) = \Delta(1, N) \geq [6\theta - 9Q_{(s-1)} - 3\tau_s^c]/[10Q_{(s-1)} - 6\theta] = [3/2][\tau_{(s-1)}^c - \tau_s^c]/[5Q_{(s-1)} - 3\theta]$.

Hence, even in the absence of an institutional constraint, firms do not desire to add capacity in the deviation period, if the discount factor (δ), the extent of irreversibility (N), or the tariff reduction ($\tau_{(s-1)}^c - \tau_s^c$) is large, or if capacity is much above the no trade agreement level.

Finally, Proposition 2 shows that if no institutional constraint exists and the deviating country does add capacity in the deviation period, then Ω_s^{id} shifts down and Ω_s^{ic} shifts up. In this case, tariff reductions still occur, however, their magnitude increases and tariffs drop at a faster rate.

Figure 1

Timing of Actions in Period t

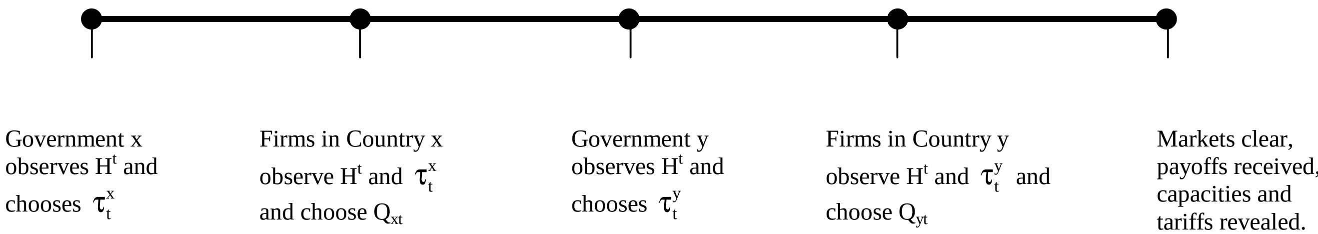

Figure 2

$\forall\ \delta > 1/3$ if $N \rightarrow \infty$

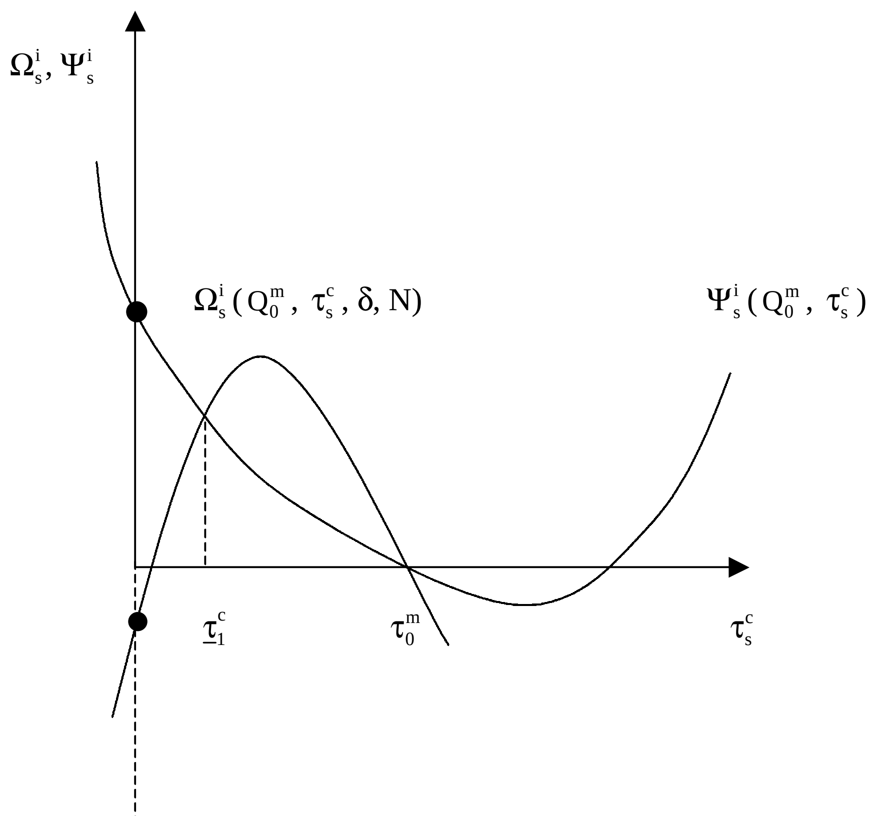

Figure 3

If $Q_{is} > Q_{i(s-1)}$

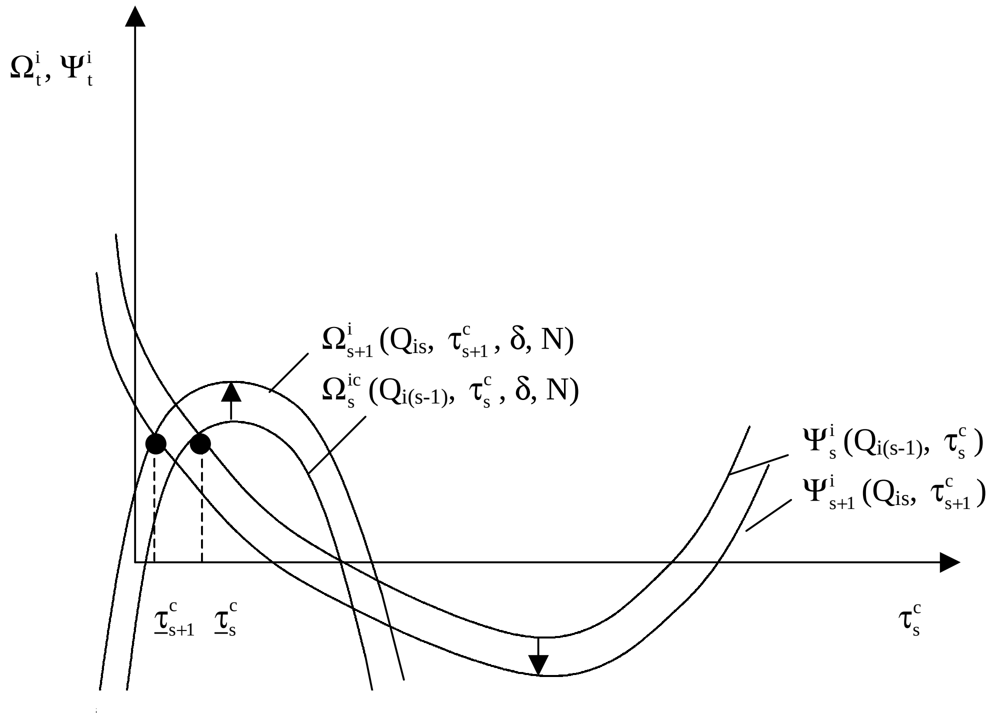

Figure 4

Deviation and Trade War Profits

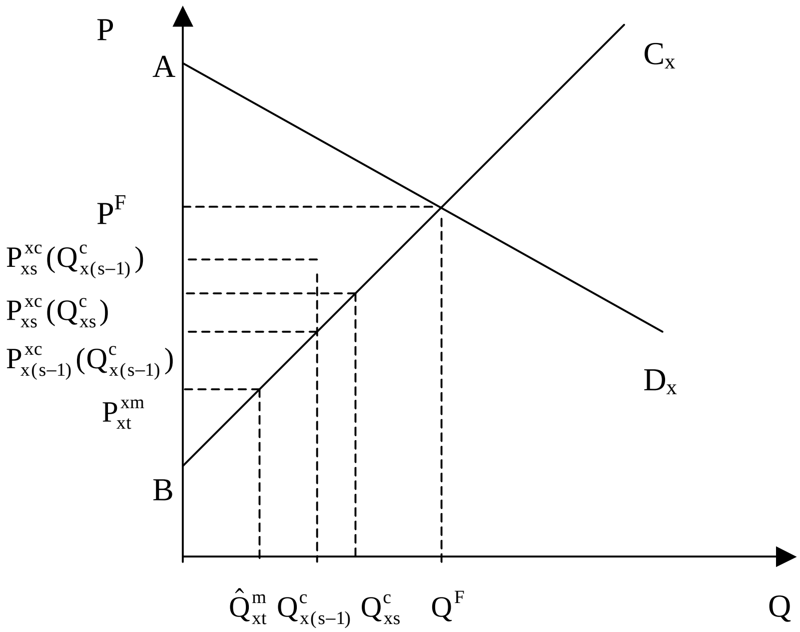

Figure 5

If $\delta > \delta^c_0(N)$

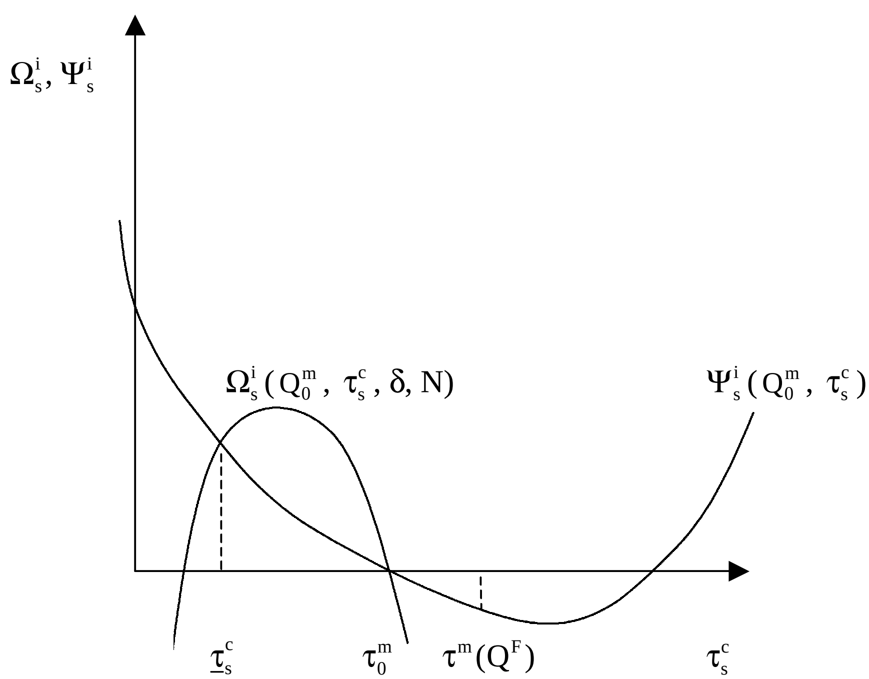

Figure 6

iff $\delta \geq \delta^F(N)$

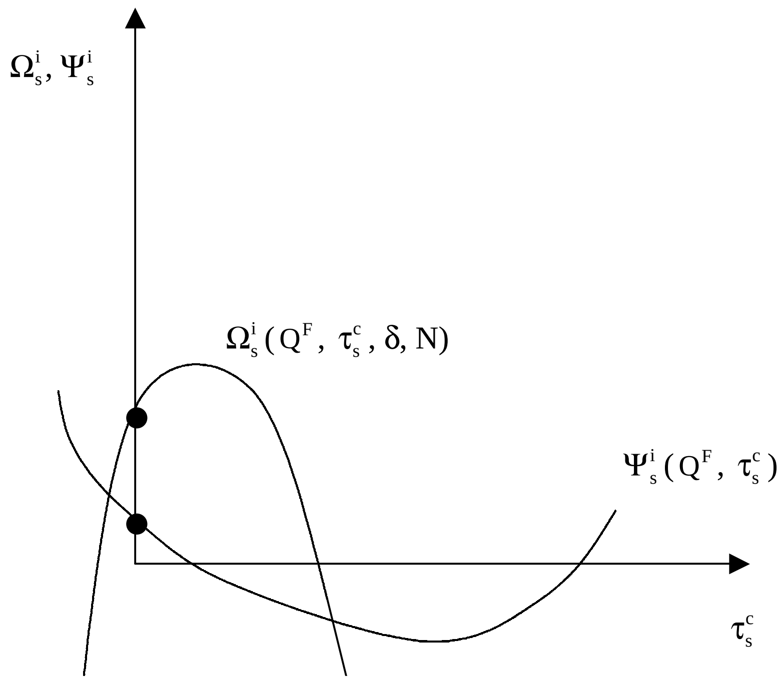

Figure 7

If N is large and $\theta' > \theta$

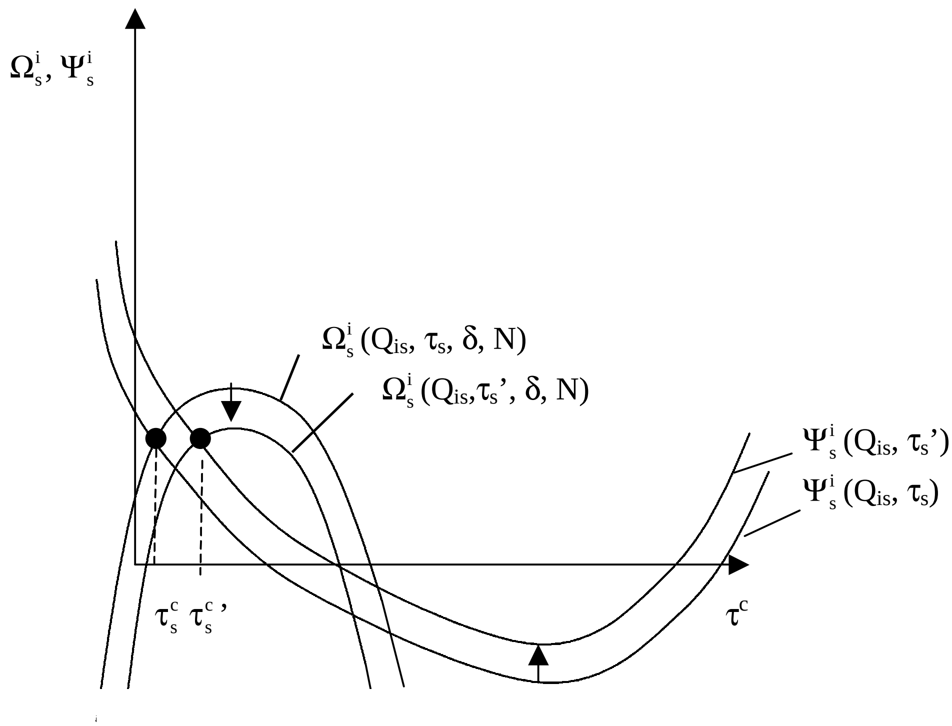

Figure 8

Evolution of the Tariff Path