

International Trade Agreements*

Giovanni Maggi

Yale University, FGV/EPGE and NBER

Abstract

In this chapter I survey the recent theoretical and empirical advances in the economics of international trade agreements, focusing on three main topics: the motives for trade agreements; the design of rules and institutions for trade policy; and regional trade agreements.

Keywords

International trade agreements

JEL classification codes

F13, F53, K33

1. INTRODUCTION

The starting point for this survey is represented by the two chapters on trade agreements in the previous volume of the *Handbook of International Economics* (1995), namely Robert Staiger's chapter "International Rules and Institutions for Trade Policy" and Richard Baldwin and Anthony Venables' chapter "Regional Economic Integration." For the most part I will focus on the post-1995 advances in the literature on trade agreements; I refer the reader to the previous volume of this handbook for the pre-1995 literature.

Before I plunge into the literature, however, it is useful to start with a quick review of the main developments that have occurred in the real world of international trade agreements since 1995. My aim is not to embark in a comprehensive discussion of these developments, but simply to provide a stylized historical context for the literature that I will survey.

The first major development is that, after the completion of the Uruguay Round in 1995, the General Agreement on Tariffs and Trade (GATT) has been replaced by the World Trade Organization (WTO). The WTO is a considerably more developed international institution than its predecessor, with a broader set of functions that include

* I am grateful to Bob Staiger and Rachel McCulloch for being my discussants at the handbook conference and providing very detailed and useful comments on an earlier draft, to co-editor Elhanan Helpman for providing an additional set of helpful suggestions, and to Kyle Bagwell, Matthew Grant, Andrés Rodríguez-Clare and participants in the handbook conference for helpful comments and discussions.

not only the implementation of the commitments made by member countries in the Uruguay Round, but also a relatively sophisticated judicial system, known as the “Dispute Settlement Procedure.”

A second important fact has been the growing role of developing and newly industrialized countries in the WTO and in regional trade agreements. An important milestone in this respect was the 2001 accession of China to the WTO. Another significant manifestation of this trend has been the growing involvement of newly industrialized countries (especially Brazil, India and, after 2001, China) in the WTO dispute settlement system.

The third major development has been a tremendous acceleration in the formation of regional trade agreements. For example, during the first 10 years of the WTO (1995–2005) the number of regional trade agreements in force more than tripled, from 58 to 188. Currently the total number of regional trade agreements in force is about 380, and several potential new ones are currently being negotiated, including one between the US and the EU, which would constitute the largest regional trade agreement in the world.

The fourth and fifth developments are perhaps better described as *non*-developments. After the completion of the Uruguay Round, the WTO member countries embarked in a new major round of multilateral negotiations, the so-called Doha Round. These negotiations started in 2001 but have stalled, and at this point it is safe to say that the Doha Round is effectively dead. Negotiations have stumbled mostly over divisions between developing and developed countries, but there has also been considerable contention between the EU and the US over agricultural subsidies. This lack of progress in multilateral trade liberalization is probably related to both of the developments I mentioned above—the proliferation of regional trade agreements and the growing role of developing countries—but the deep reasons for the failure of the Doha Round are an open question.

The second non-development is—contrary to the one discussed just above—a positive one: the *existing* rules and institutions have held up rather well, even in the face of challenges such as the 2008 Great Recession, which led to a dramatic (though temporary) collapse of trade flows, and the accession of China to the WTO. In particular, the WTO’s dispute settlement system has been quite effective. Indeed, many scholars have argued that the enforcement and judicial aspects of the WTO are stronger now than during the GATT years. Furthermore, if judged by the standards of international organizations, it is safe to say that the WTO has established itself as one of the most, if not the most, successful international organization in terms of enforcement and dispute settlement.

This concludes my stylized portrait of the recent developments in trade agreements, and against this backdrop I now turn to my discussion of the recent advances in the academic literature on trade agreements.

Section 2 focuses on theoretical and empirical research investigating the purpose of trade agreements. In the theoretical part, I focus mostly on three theories: the “terms of trade” theory, according to which the purpose of a trade agreement is to prevent governments from manipulating terms of trade; the “domestic commitment” theory, according to which a trade agreement can provide a government with a means to tie its own hands

vis-à-vis domestic agents; and the “New Trade” theory, which emphasizes the role that a trade agreement can play in the presence of imperfect competition. The empirical part of the discussion focuses on two sets of contributions: those aimed at testing the predictions of the main theories, and those that study the impacts of trade agreements on trade barriers and trade flows in a more descriptive way.

Section 3 discusses research aimed at understanding the design of trade agreements, with a particular emphasis on the role of transaction costs. In particular, I will focus on two distinct types of transaction cost: *contracting frictions* and *enforcement frictions*. I will argue that taking these transaction costs into account is essential for understanding the design of substantive policy rules (such as tariff ceilings, non-discrimination rules, etc.), enforcement rules (which regulate punishment/retaliation), and dispute settlement procedures.

Section 4 discusses recent research on regional trade agreements. In particular, I will focus on the economic and political determinants of regional trade agreements; on the impacts of such agreements on its members’ external trade barriers and on multilateral trade liberalization; and on the design of rules for trade negotiations, that is, on whether the formation of regional trade agreements should be subject to constraints or even prohibited.

Finally, Section 5 offers some concluding remarks and some thoughts about possible avenues for future research.

2. THE MOTIVES FOR TRADE AGREEMENTS

The most basic question regarding trade agreements (TAs) is why countries sign them in the first place. In spite of this being the most fundamental of questions, it is only in the last 15 years or so that the academic literature has made substantial progress in answering it.

Krugman (1997, pp. 113–120) made a famous pessimistic statement: “*Anyone who has tried to make sense of international trade negotiations eventually realizes that they can only be understood by realizing that they are a game scored according to mercantilist rules. (...) The implicit mercantilist theory that underlies trade negotiations does not make sense on any level, indeed is inconsistent with simple adding-up constraints; but it nonetheless governs actual policy (...) the economic theory underlying trade negotiations is nonsense.*” The last 15 years of research on TAs are in some way an attempt to prove Krugman wrong, and argue instead that the logic of economics (in a broad sense that includes also the logic of political economy) can to a large extent make sense of real-world TAs.

In this section I will offer my critical survey of the main theories for why countries sign TAs and of the small but growing empirical literature on this subject.

Before I proceed, I need to make clear what is the organizing principle of this section. The distinction between analyzing the motives for a TA (which is the subject of this section) and analyzing the design of a TA (which is the subject of the next section) can

sometimes be blurred, since the two aspects are obviously inter-related. But I think it is important to keep these two aspects conceptually distinct. In the present section, I abstract from issues of institutional/contract design (which, as I will argue below, are interesting only in the presence of transaction costs) by assuming that governments can negotiate directly and costlessly over all policies (and, in the presence of uncertainty, over all contingencies).

There are two broad stories for why governments sign TAs. The first one is that a TA can provide governments with an escape from an international Prisoners' Dilemma, which is in turn caused by international externalities from trade policy: these include the classic terms-of-trade externalities and the "New Trade" externalities that arise with imperfect competition (such as "delocation" and "profit-shifting" externalities). The second broad story for why governments sign TAs is that these may provide governments with a commitment device vis-à-vis domestic actors, such as industrial lobbies or individuals making investment decisions. I will start by focusing on the classic terms-of-trade story.

2.1. The Terms-of-Trade Theory

The type of international externality from trade policy that has received by far the most attention in the literature (dating back at least to [Johnson, 1953](#)) is the terms-of-trade (TOT) externality that arises in a perfectly competitive environment. As it will become clear in [Section 2.3](#), TOT externalities play an important role also with imperfect competition, but their role is clearest in the case of perfect competition, so I will focus on this case here.

I will proceed in two steps. I will first illustrate the TOT theory using a simple "workhorse" model that is very structured and delivers simple formulas and a number of specific predictions. This model is convenient also because it can be used as a basis to illustrate a simple version of the domestic-commitment theory, as I will show in [Section 2.2](#). I will then present a more general version of the TOT theory, which has been developed mostly by Bagwell and Staiger.

2.1.1. A Simple Workhorse Model

My workhorse model is essentially a simplified version of [Grossman and Helpman's \(1995a\)](#) "Trade Wars and Trade Talks" model, where governments may have two motives for unilateral trade policy intervention, namely a terms-of-trade motive and a political-economy motive.

Consider a world with two countries (H and F) and with three sectors, a numeraire sector (0) and two non-numeraire sectors ($j = 1, 2$). All citizens have the same utility function, which takes the form $U = c_0 + \sum_{j=1}^2 u_j(c_j)$, where each u_j is increasing and concave. The numeraire good is produced one-for-one from labor, so the wage is pinned down to one, while good j is produced from labor and capital according to the constant-returns production function $y_j = F_j(k_j, l_j)$.

Labor is perfectly mobile across sectors. Capital for the moment is assumed to be immobile across sectors (so that it can be effectively viewed as a specific factor); in Section 2.2, where I consider the domestic-commitment theory, I will assume that capital is immobile in the short run but can move across sectors in the long run.

The owners of capital represent a negligible fraction of the total population; this will simplify the political-economy environment. The size of the population is equal to one in each country. Country H is the natural importer of good 1 and country F of good 2.

Assume the numeraire good is freely traded. Each government can choose specific trade taxes/subsidies in the non-numeraire sectors. Later I will consider domestic policies as well.

Next I describe the political-economy environment. In each country, the owners of capital in a given sector may or may not be organized into a lobby. The government's objective is $\Omega^G = aW + C$, where W is aggregate welfare and C denotes total contributions received from lobbies. If sector j is organized, the lobby's objective is $\Omega_j^L = \Pi_j - C_j$, where Π_j denotes returns to capital in sector j . Analogous notation (but with asterisks) will apply to the F country.¹

Given the quasi-linear preferences, the specific-factor structure, and the wage pinned down at one, this is essentially a partial-equilibrium setting. Defining welfare as aggregate indirect utility, Home welfare can be written (using standard techniques) as $W = Y + S + R$, where Y is total factor income, S is total consumer surplus (summed over the two goods) and R is revenue from trade policy (positive or negative). An analogous expression holds for Foreign welfare.

For each sector j , the demand functions in the two countries are denoted by $d_j(p_j)$ and $d_j^*(p_j^*)$, the supply functions by $\gamma_j(p_j)$ and $\gamma_j^*(p_j^*)$, and the import demand functions by $m_j(p_j)$ and $m_j^*(p_j^*)$.

Let τ_j (τ_j^*) denote the specific trade tax/subsidy chosen by Home (Foreign) in sector j . If good j is imported by Home, τ_j is interpreted as an import tax, and τ_j^* as an export subsidy, and vice-versa if the good is exported by Home. The price arbitrage condition ($p_j = p_j^* + \tau_j - \tau_j^*$) and the market clearing condition ($m_j(p_j) + m_j^*(p_j^*) = 0$) together determine equilibrium prices as functions of trade policies: $p_j(\tau_j - \tau_j^*)$ and $p_j^*(\tau_j - \tau_j^*)$. Finally, define the "world price" of good j as $p_j^W = p_j - \tau_j = p_j^* - \tau_j^*$.

I start by focusing on the noncooperative scenario. In each country, assume that government and lobbies choose trade policy (and contributions) by Nash bargaining, taking foreign trade policy as given. Given that bargaining is efficient, trade policy in each country maximizes the joint surplus of government and lobbies given foreign trade policy. Thus Home trade policy maximizes $\Omega \equiv \Omega^G + \sum_{j=1}^2 I_j \Omega_j^L = aW + \sum_{j=1}^2 I_j \Pi_j$,

¹ I note here that very little research has been done on the role of *informational* lobbying (as opposed to *quid pro quo* lobbying) in influencing trade negotiations. The only paper of this kind that I am aware of is Milner and Rosendorff (1996). Other papers on this general theme are Tovar (2011) and Ludema et al. (2010), but these papers focus only on unilateral trade policy choices, not trade agreements.

where $I_j = 1$ if sector j is organized and zero otherwise. Similarly, Foreign trade policy maximizes $\Omega^* \equiv a^* W^* + \sum_{j=1}^2 I_j^* \Pi_j^*$.

As mentioned above, this model can be viewed as a simplified version of [Grossman and Helpman \(1995a\)](#).² Note also that this model is equivalent to one where each government maximizes a politically-adjusted welfare function that attaches an extra weight to the organized sectors' rents, as for example in [Baldwin \(1987\)](#).

Maximizing Ω with respect to τ_j and Ω^* with respect to τ_j^* yields the following formulas for τ_j and τ_j^* :

$$\tau_j = \frac{1}{\eta_j^*} + \frac{I_j}{a \cdot \eta_j \cdot \frac{m_j}{\gamma_j}}, \quad (1)$$

$$\tau_j^* = \frac{1}{\eta_j} + \frac{I_j^*}{a^* \cdot \eta_j^* \cdot \frac{m_j^*}{\gamma_j^*}}, \quad (2)$$

where $\eta_j^* \equiv \frac{|m_j^{*'}|}{m_j^*}$ and $\eta_j \equiv \frac{|m_j'|}{m_j}$. The first term in each formula is the well-known Johnson's optimum trade tax, which captures the *terms of trade* motive for trade policy; the second part of each formula captures the *political* motive for trade policy. Note that the two motives go in the same direction for organized import-competing industries (both call for a tariff), but are in conflict for organized export sectors (TOT considerations call for an export tax, political considerations call for an export subsidy).

We are now ready to examine the trade policies that are selected if the two countries negotiate a TA. Assuming that the TA maximizes the joint surplus of all governments and lobbies, that is $\Omega + \Omega^*$,³ we obtain:

$$\tau_j - \tau_j^* = \frac{I_j}{a \cdot \eta_j \cdot \frac{m_j}{\gamma_j}} - \frac{I_j^*}{a^* \cdot \eta_j^* \cdot \frac{m_j^*}{\gamma_j^*}}. \quad (3)$$

Note that for each sector j the optimal agreement pins down only the net trade tax $\tau_j - \tau_j^*$, not the exact levels of τ_j and τ_j^* . To understand intuitively why, consider the case in which governments maximize welfare: $a = a^* = \infty$. Then the formula above yields $\tau_j = \tau_j^*$. A tariff of \$1 and an export subsidy of \$1 on the same good neutralize

² One simplification relative to [Grossman and Helpman \(1995a\)](#) is that the share of the population represented by lobbies (the α_L parameter) is negligible. Another simplification is that the interaction between government and lobbies is modeled as a Nash bargaining game, whereas in Grossman and Helpman it is modeled as a common-agency game, but in both cases a country's trade taxes maximize the joint surplus of government and lobbies.

³ This will be the case, for example, if trade negotiations take the form of a Nash bargain that involves the governments and all lobbies and international transfers are available. I note that [Grossman and Helpman \(1995a\)](#) assume a two-stage game where lobbies first offer contribution schedules to their respective governments, and then governments bargain in Rubinstein fashion, but the policy outcome in their model is the same as in the simplified model I consider here. Also note that, even if international transfers are not available, such transfers can be indirectly effected by adjusting import and export tax levels, as I explain below.

each other's effect on domestic prices, and the only effect that remains is a revenue transfer from the exporting country to the importing country; the condition $\tau_j = \tau_j^*$ is the partial-equilibrium analog of the well-known "Mayer curve."⁴

If political motivations are present in the governments' objectives (a and a^* are less than infinity), the efficient policies will reflect only these motivations, and not the TOT motivations: *an efficient agreement simply removes TOT considerations from trade policies*. This in turn suggests that the only source of inefficiency in the noncooperative equilibrium is the governments' temptation to manipulate TOT. Or put another way, even if governments are politically motivated, the reason for signing a TA is inherently economic, not political. As I discuss next, this basic message is further developed and generalized by Bagwell and Staiger.

2.1.2. The Bagwell-Staiger Approach

In an influential series of papers, Bagwell and Staiger have argued that the TOT theory—and in particular the conclusion that the only motive for a TA can be traced to the governments' temptation to manipulate TOT—is considerably more general than previously thought. I refer the reader to Bagwell and Staiger's (2002) book and Bagwell and Staiger's (2010a) survey for more detailed expositions of their work. In this section I will offer a succinct exposition of Bagwell and Staiger's theory under perfect competition; I will discuss their models with imperfect competition in the next section.

Consider a two-country, two-good, perfectly competitive world, with the good exported by Home taken as the numeraire. The production possibility frontier is assumed to be concave, and the goods are assumed to be normal in consumption. The model allows for domestic distortions, such as consumption or production externalities, that call for corrective domestic policies (but monopoly distortions are not allowed).⁵ Each government has access to a complete set of (trade and domestic) policy instruments. It is useful to partition policies in two categories: (i) Tax instruments that create wedges between local prices and the world price, or more specifically, trade taxes, production taxes, and consumption taxes. I will refer to these as "wedge policies"⁶; (ii) Other policies (e.g. labor standards) that may affect market conditions but cannot affect price wedges. I will refer to these as "non-wedge policies."

Key to Bagwell and Staiger's approach is the way government objectives are represented. These are represented in reduced form as $\Omega(v, p, q, p^w)$ and $\Omega^*(v^*, p^*, q^*, p^w)$, where p (p^*) is the Home (Foreign) consumer price, q (q^*) the Home (Foreign) producer

⁴ See Mayer (1981), who characterizes the locus of efficient tariff combinations in a general equilibrium model with two goods and two countries.

⁵ The framework I present here can be seen as combining three variants of the Bagwell-Staiger model: Bagwell and Staiger (1999a), which considers only trade taxes; Bagwell and Staiger (2001), which considers trade taxes and domestic standards; and Bagwell and Staiger (2006), which considers trade taxes and production subsidies.

⁶ Of course there is a degree of redundancy in these three taxation instruments, since a trade tax is equivalent to a combination of a production subsidy and a consumption tax.

price, v (v^*) the vector of Home (Foreign) non-wedge policies, and p^w the world price. Since wedge policies can be written as price wedges, they need not be included as arguments in the objective functions. Of course, the equilibrium price levels in general will depend on all policies.

A central feature of this setting is that a government's policies affect the other government's payoff through a unique channel: the world price. In other words, the only international externality is the TOT externality. An implicit assumption in this setting is that there are no non-pecuniary international externalities (such as cross-border pollution). But aside from this restriction, the above representation of governments' objectives is general enough to capture the presence of political-economy motives for protection; indeed, Bagwell and Staiger argue that many of the existing political-economy models (including the Grossman–Helpman model presented above) can be represented in this fashion.

The only structure Bagwell and Staiger impose on government preferences is that, for given domestic prices, a government dislikes a worsening of terms of trade: $\frac{\partial \Omega}{\partial p^w} < 0 < \frac{\partial \Omega^*}{\partial p^w}$. This seems like a reasonable restriction, since domestic interest groups care about domestic prices, not directly about world prices. If p^w increases while domestic prices are kept constant, there is simply a transfer of revenue from Home to Foreign (given the assumption of normal goods), so this amounts to assuming that, all else equal, a government values revenue. Finally, the Lerner paradox and the Metzler paradox are assumed away.⁷

As usual, the analysis starts with the noncooperative scenario, that is the Nash equilibrium of the game where governments simultaneously choose policies. Three basic points can be made in this setting.

The first point is that any Nash equilibrium is Pareto-inefficient from the point of view of the governments' objectives. This result is intuitive, since a country's policies exert externalities on the other country through TOT, and hence unilateral policy choices will generically not be efficient.

The second point is that the governments' temptation to manipulate TOT is the only source of inefficiency in the Nash equilibrium, and hence it is the only motive for a TA. Bagwell and Staiger establish this point by considering a diagnostic test to ascertain if TOT manipulation is the only cause of the "disease" in the noncooperative equilibrium. The test is based on the following question: if each government did not value the pure terms-of-trade effects of its policies, would governments make efficient choices? If the test is positive, the diagnosis is that TOT manipulation is the only cause of the disease.

More specifically, Bagwell and Staiger define the *politically optimal* (PO) policies as those that would result if governments did not value changes in p^w . If the PO policies are efficient, then TOT manipulation is deemed to be the only cause of inefficiency in the noncooperative equilibrium. In the setting under consideration, Bagwell and Staiger

⁷ The Lerner paradox occurs when an increase in a country's tariff leads to an increase in the world relative price of the imported good; the Metzler paradox occurs when an increase in a country's tariff leads to a decrease in the domestic relative price of the imported good.

show that the PO policies are indeed efficient. A simple intuition for this result can be gained by considering a setting where utility is transferrable, with changes in p^w acting as pure transfers, so that efficient policies must maximize the global payoff $\Omega + \Omega^*$. In this case, at a political optimum, Home's policies maximize Ω given p^w and Foreign policies maximize Ω^* given p^w , therefore the global payoff must be maximum because p^w is a pure transfer.

The third point is that trade volume at the Nash equilibrium is inefficiently low, and a mutually beneficial TA must entail a reciprocal expansion of market access relative to the Nash equilibrium.⁸ I will provide a simple local intuition for this result, abstracting from domestic policies for simplicity. Starting from the Nash equilibrium, a small increase in a trade tax has a negative externality on the trading partner. This is not self-evident, since an increase in a country's trade tax could in principle have a positive effect on the trading partner through the latter's local price. But this cannot be the case locally at the Nash equilibrium: the optimality of a country's unilateral policies implies that, at the Nash equilibrium, any effect through the local price cannot outweigh the adverse effect through the world price, so the externality from an increase in the foreign trade tax is negative.⁹ Given the negative international externalities from trade taxes, it is intuitive that, starting from the Nash equilibrium, the only way to achieve a Pareto improvement is to decrease both trade taxes, which in turn will expand trade.

In light of the three points highlighted above, Bagwell and Staiger conclude that under the assumptions of their model the only purpose of a TA is to preclude countries from manipulating TOT, and this in turn entails a reciprocal expansion of market access relative to the noncooperative equilibrium. This prediction of the model resonates with the emphasis placed by the GATT-WTO on the exchange of market access between countries.

Next I highlight a prediction that presents a special challenge for the TOT theory, because it is at odds with observed TAs. According to the TOT theory, *a TA should tend to increase export subsidies* relative to the noncooperative equilibrium, whereas in reality export subsidies are typically restricted by TAs. This counterfactual prediction of the TOT

⁸ Bagwell and Staiger make a distinction between an expansion of "market access" and an expansion of trade volume (see Bagwell and Staiger, 2001, pp. 537–538). The former is a weaker condition than the latter: a policy change is said to expand market access to country A if it shifts out country A's import demand curve for at least *some* world price. As Bagwell and Staiger show, a mutually beneficial trade agreement must entail a reciprocal expansion of market access, but in general it need not entail an expansion of import volumes; it will entail an expansion of import volumes under the additional assumption that any policy change shifts a country's import demand curve in the same direction for all world-price levels. In my intuitive discussion in the text I abstract from domestic policies, so this distinction is not necessary, and a mutually beneficial agreement always entails an expansion of trade volumes.

⁹ To see this formally, note first that if only trade taxes are available, we can write Home's payoff as $\Omega(p, p^w)$. Letting τ and τ^* denote trade taxes, we can write the externality of τ^* on Home (with a slight abuse of notation) as $\Omega_{\tau^*} = \Omega_p p_{\tau^*} + \Omega_{p^w} p_{\tau^*}^w = (\Omega_p + \Omega_{p^w}) p_{\tau^*}^w$ (where I used $p_{\tau^*} = p_{\tau^*}^w$). At a Nash equilibrium, the first order condition (FOC) is $\Omega_{\tau} = \Omega_p p_{\tau} + \Omega_{p^w} p_{\tau}^w = 0$, which using $p_{\tau} = p_{\tau}^w + 1$ yields $\Omega_p = -\frac{\Omega_{p^w} p_{\tau}^w}{p_{\tau}^w + 1}$, which is negative by the no-Lerner-paradox and no-Metzler-paradox assumptions. This in turn implies $\Omega_{\tau^*} < 0$.

model can be illustrated by focusing on a model with two goods. The well-known Lerner symmetry theorem states that an import tax is equivalent to an export tax, thus we can suppose without loss of generality that each government uses only an export tax (or if negative, an export subsidy). In this case, if the noncooperative equilibrium entails export subsidies (which is possible if export interests are politically strong), it is easy to show that a mutually beneficial TA must increase their levels. Intuitively, increasing a country's export subsidy has a positive TOT externality on the other country, so governments "under-subsidize" exports in equilibrium. I will refer to this feature as the *export subsidy puzzle* in the TOT theory. As I discuss below, possible ways to resolve this puzzle include considering domestic-commitment motives and "New Trade" motives for TAs.

Thus far I have focused on a two-country world. Extending the analysis to a multi-country world introduces new considerations. As Bagwell and Staiger (1999a) make clear, when trade policies can discriminate across trading partners, there is no longer a single world price but a whole vector of bilateral world prices, and importantly, international externalities can no longer be viewed as traveling solely through world prices. As a consequence, if trade policies can be discriminatory, the PO policies are inefficient. To understand this point, focus on the impact of foreign trade policies on the Home country. Define the multilateral TOT as an import-weighted average of bilateral world prices. Since the import weights depend on foreign local prices, now international externalities travel not only through world prices but also through foreign local prices. It is then intuitive that the PO policies are not efficient. On the other hand, if governments are constrained by a Most Favored Nation (MFN) rule to choose non-discriminatory trade policies, then Bagwell and Staiger show that all international externalities are channeled through a single world price, and again the PO policies are efficient. To summarize, the PO policies are efficient if and only if trade policies are constrained by the MFN rule.

The result I just highlighted can be interpreted in more than one way. Bagwell and Staiger argue that the result confirms the general point that in a perfectly competitive environment the only purpose of a TA is to prevent the manipulation of TOT. But one could argue that the appropriate thought experiment should diagnose the cause of the disease in a scenario where *no* institutional constraints are in place, not even the MFN rule, in which case PO policies are inefficient and one should conclude that TOT manipulation is not the only motivation for a TA. Thus there is a legitimate question as to which of the two diagnostic tests (with unconstrained policies or with MFN-constrained policies) is more informative about the deep motivation for a TA.

Bagwell and Staiger build on the model outlined above to argue that it can explain some key rules of the GATT-WTO, such as reciprocity, MFN, and the so-called "non violation" rule. I will come back to these themes in the next section, where I focus on the design of TAs, but here I wish to re-iterate a point already mentioned above: in a world without transaction costs, the theory would not be able to explain any such rules, because then governments could simply negotiate directly on the policy levels, and there would be

no need for additional rules, so this second part of Bagwell and Staiger's theory implicitly relies on the presence of some kind of transaction costs. As already mentioned, in the present section I am assuming that there are no transaction costs, and hence governments can negotiate directly and costlessly over all policies, so I postpone issues of rules design to the next section.

2.2. The Domestic-Commitment Theory

The TOT theory is by far the one with the deepest roots in the literature, but it is not clear that TOT considerations are the whole story behind TAs, for at least two reasons. First, casual empiricism suggests that small countries (which have negligible influence on world prices) often agree to significant cuts in their trade barriers when they join a TA, an observation that is not easy to reconcile with the TOT theory.¹⁰ And second, as I mentioned above, the TOT theory implies that TAs should tend to increase export subsidies relative to their noncooperative levels, which is a counterfactual prediction. An alternative theory that can explain these observations is based on the idea that a TA can help a government tie its own hands vis-à-vis domestic actors.¹¹

There are several models in the literature that fall within the broadly defined domestic-commitment theory. Some are of a purely economic nature, for example [Staiger and Tabellini \(1987\)](#), [Tornell \(1991\)](#), and [Lapan \(1988\)](#), and some are of a political-economy nature, in particular [Maggi and Rodriguez-Clare \(1998, 2007\)](#), [Mitra \(2002\)](#), [Brou and Ruta \(2009\)](#), [Limão and Tovar \(2011\)](#), and [Liu and Ornelas \(2012\)](#). Since the former type of domestic-commitment models was covered by [Staiger's \(1995a\)](#) chapter, I will focus on the latter type, and in particular on the version due to [Maggi and Rodriguez-Clare \(1998, 2007\)](#).

The general idea proposed by Maggi and Rodriguez-Clare is that a TA can serve as a commitment device for a government to close the door to domestic lobbies. It has been argued by a number of scholars and commentators that this type of motivation was central to Mexico's negotiations of the North American Free Trade Agreement (NAFTA). For example, [Whalley \(1998\)](#) argued that Mexican negotiators of NAFTA "were less concerned to secure an exchange of concessions between them and their negotiating partners, and were more concerned to make unilateral concessions to larger negotiating partners with whom they had little negotiating leverage... The idea was clearly

¹⁰ The reason I use the expression "casual empiricism" is the following. There is little doubt that at least in *some* cases countries with negligible monopsony power on given goods have agreed to significant tariff cuts on those goods, but I am not aware of any empirical study that investigates whether this is the case more systematically.

¹¹ Interestingly, in the same 1997 essay where Krugman declared it impossible to understand trade negotiations from a rational perspective, he left a small opening for the domestic-commitment theory of trade agreements, although still with some degree of skepticism. He summarizes this theory as maintaining that "the true purpose of international negotiations is arguably not to protect us from unfair foreign competition, but to protect us from ourselves," then states that "one cannot dismiss such political-economy arguments as foolish," but questions whether in reality international agreements are truly effective in achieving this purpose.

to help lock in domestic policy reform.” Similarly, [Bajona and Chu \(2010\)](#) view China’s accession to WTO as a way to “... lock-in the agenda for fundamental domestic reforms, which has been difficult to implement by domestic measures alone.”

Notice however that, if one considers the typical models of lobbying that have been proposed in the literature, in particular those in the tradition of [Grossman and Helpman’s \(1994\)](#) “Protection for Sale,” it is not clear why a government would ever want to tie its own hands, since it derives positive rents from the political process.

[Maggi and Rodriguez-Clare \(1998\)](#) provide a theoretical justification for the domestic-commitment argument based on a simple dynamic model. The idea is that a government can derive rents from the interaction with lobbies in the short run, but in the long run this will distort the allocation of resources, because investors will overinvest in the sectors that are expected to get trade protection, and the government is not compensated for this long-run distortion. As a consequence, the government may be better off committing to free trade ex-ante, thereby shutting down the lobbying process.¹²

The basic points can be illustrated within the workhorse model of Section 2.1. Consider the same economic and political structure as in that model, but now suppose that H is a small country, while F is a large “rest of the world.” Also assume for simplicity that both sectors 1 and 2 are politically organized.

Consider the following timing: (0) the small-country government chooses whether to commit to free trade; (1) capital is allocated; and (2) given the capital allocation, trade policy and contributions are determined by Nash bargaining between the government and the lobbies (with σ denoting the government’s bargaining power). This timing captures the idea that capital is mobile in the long run but not in the short run.

Suppose first that the government does not commit to free trade. Let us proceed by backward induction and find the second-stage equilibrium payoffs given the capital allocation. For the Home country, let K denote the vector of capital allocations and τ the vector of trade policies. Also, let $W(\tau, K)$ and $\Pi(\tau, K)$ denote respectively the levels of general welfare and the aggregate returns to capital in sectors 1 and 2 as functions of trade policies and capital allocations. Given that the government and the lobbies engage in Nash bargaining over policies and contributions, the first step is to derive the status-quo (disagreement) payoffs. In the status quo, lobbies give no contributions and the government chooses the welfare-maximizing policy, which is free trade, hence the government’s status-quo payoff is $aW(0, K)$, and the lobbies’ total status-quo payoff is $\Pi(0, K)$. The next step is to write down the joint surplus of the government and the lobbies:

$$J(K) = \max_{\tau} [aW(\tau, K) + \Pi(\tau, K)] - [aW(0, K) + \Pi(0, K)].$$

¹² I note that, while Maggi and Rodriguez-Clare focus on a setting where a government can be pressured only by its domestic lobbies, similar benefits from committing to free trade may arise if a government can be influenced also by foreign lobbies. For a paper that documents the empirical importance of foreign lobbying, see [Gawande et al. \(2006\)](#).

The government walks away with a share σ of this joint surplus, therefore its payoff in the second stage is given by $aW(0, K) + \sigma J(K)$.

The next step is to derive the equilibrium allocation in the first stage, which I denote \hat{K} . The key point is that, if $\sigma < 1$, this will generically be different from the free trade allocation ($\hat{K} \neq K^{FT}$), and hence inefficient, while $\hat{K} = K^{FT}$ if $\sigma = 1$. This is intuitive, because as long as lobbies have any bargaining power ($\sigma < 1$), the presence of lobbying distorts the net returns to capital relative to free trade. If, on the other hand, lobbies have no bargaining power ($\sigma = 1$), they will walk away from the bargain with no surplus, and hence the lobbying process does not affect the returns to capital *net of contributions*, so the equilibrium allocation is efficient. With this in mind, we can write the government's equilibrium payoff in the no-commitment scenario as $G^{NO} = aW(0, \hat{K}) + \sigma J(\hat{K})$.

Now suppose the government commits to free trade. In this case, expecting free trade, capital owners will make efficient allocation decisions: $K = K^{FT}$, and hence the government's payoff in this case is $G^{COMM} = aW(0, K^{FT})$.

The government will commit to free trade if and only if $G^{COMM} > G^{NO}$. Now observe that: (i) if $\sigma = 0$, then $G^{COMM} > G^{NO}$, because $W(0, K^{FT}) > W(0, \hat{K})$; and (ii) if $\sigma = 1$, as I noted above we have $\hat{K} = K^{FT}$, and since $J(K^{FT}) > 0$ then $G^{COMM} < G^{NO}$. We can then conclude that if σ is sufficiently low the government will commit to free trade, and if σ is sufficiently high it will not. Moreover, under some conditions G^{NO} will be increasing in σ , in which case there will be a critical level of σ below which the government commits to free trade and above which it does not. Thus the model yields an interesting prediction: countries whose governments have a weaker bargaining position vis-à-vis domestic lobbies should be more likely to join a TA.

Another prediction generated by the model concerns the impact of the parameter a , the government's valuation of welfare relative to contributions. Provided σ is sufficiently small, the value of commitment ($V = G^{COMM} - G^{NO}$) is non-monotonic in a : it starts negative, then it turns positive, and eventually it approaches zero as $a \rightarrow \infty$.¹³ This in turn implies that, if there is a small cost of joining the agreement, the government will choose to join if a falls in some intermediate range.

Importantly, note that if export interests are organized the noncooperative equilibrium will entail export subsidies, so in Maggi and Rodriguez-Clare (1998) the government may want to commit to the elimination of export subsidies. Thus the model suggests a possible solution to the "export subsidy puzzle" highlighted above in the context of the TOT theory: if TAs are motivated by domestic-commitment issues, they will reduce export subsidies relative to their noncooperative levels.

¹³ To see this, note that (i) if $a = 0$, the government does not care about welfare, so clearly $V < 0$; (ii) if $a = \infty$, the government only cares about welfare, so tariffs are zero in the political equilibrium, hence $V = 0$; (iii) that V must be positive for a range of a if σ is sufficiently small follows from the observation made above that, for fixed $a > 0$, if σ is sufficiently small then $V > 0$.

Next I make a point that will be useful to keep in mind when I focus on the implications of incomplete contracting for TAs (Section 3). Recall that in [Maggi and Rodriguez-Clare \(1998\)](#) the inefficiency in the noncooperative equilibrium stems from the government's lack of commitment vis-à-vis domestic agents, and the core of the problem is that the government does not get compensated for the long-run distortions from trade protection. But note that the same problem can be viewed also as a problem of *incomplete contracting* between the government and domestic agents: if the government could sign a long-term contract with all the future beneficiaries of protection, in which it commits to future trade policies and gets compensated for them, the problem would disappear. Of course, if capital is mobile in the long run, this long-term contracting would have to involve all capital owners in the economy, not only those that are currently in the organized sectors, thus it seems reasonable to assume that such long-term contracting is not feasible.

[Maggi and Rodriguez-Clare \(2007\)](#) extends the previous model in four directions. First, it allows for two large countries; thus the model nests two motives for a TA: a domestic-commitment motive and a TOT motive. Second, governments can commit to arbitrary tariff levels (as opposed to free trade or nothing); moreover, they can do so through exact tariff commitments (a complete contract) or through tariff caps (an incomplete contract). Third, specific-factor owners can lobby ex-ante to influence the shape of the agreement, not only ex-post. And fourth, the model allows for different degrees of capital mobility across sectors.

The model considers the following dynamic scenario. The world is sitting at the non-cooperative equilibrium—with its associated allocation distortions—when the opportunity to negotiate an agreement arrives.¹⁴ The agreement maximizes the joint surplus of governments and lobbies. After the agreement is signed, each investor gets a chance to move his or her capital with an (exogenous) probability z . The parameter z thus captures the degree of mobility of capital. After the reallocation of capital has taken place, tariffs are chosen in each country by the government and the lobby subject to the constraints set by the agreement. Of course, this ex-post lobbying process is relevant only if the agreement leaves some discretion, that is, if the TA takes the form of tariff ceilings.

The key results of the model are four. First, the extent of trade liberalization (the tariff cuts enacted by the TA) is increasing in the degree of capital mobility (z). Intuitively, if z is higher, current lobby members care less about future protection, and hence they are less resistant to tariff cuts. This in turn suggests a further prediction, beyond those highlighted above in the context of the small-country model: tariff cuts should be deeper in sectors where capital is more mobile. This prediction seems consistent with the anecdotal

¹⁴ In the basic version of the model the opportunity to sign a trade agreement is a surprise to investors, but [Maggi and Rodriguez-Clare \(2007\)](#) also consider a version of the model in which the trade agreement is perfectly anticipated by investors.

observation that in reality trade liberalization has been hard to come by in the agricultural sector, but it would be interesting to test this prediction in a more systematic way.

The second result concerns the impact of “politics”—captured inversely by the governments’ valuation of welfare (a)—on the extent of trade liberalization: tariff cuts are deeper when politics are more important, provided the domestic-commitment motive is strong enough (z sufficiently high). This result stands in interesting contrast with the prediction of the pure TOT model, where tariff cuts if anything tend to be *less* deep when a is lower: the reason is that a lower a implies higher noncooperative tariffs, hence a lower trade volume and a weaker TOT externality, and this calls for smaller tariff cuts. Also in [Maggi and Rodriguez-Clare \(2007\)](#), a lower a implies higher noncooperative tariffs, but this in turn implies a bigger allocation distortion, and hence bigger tariff cuts are called for. If z is high, this consideration dominates the previous one.

At a more fundamental level, the divergence in results highlighted above is a manifestation of a key difference between the domestic-commitment theory and the TOT theory. In the domestic-commitment theory, the motive for a TA is inherently political, since the TA is directly aimed at blunting domestic lobbying pressures, thus the TA is *directly* affected by political parameters such as the governments’ valuation of welfare; whereas in the TOT theory, the motive for a TA is inherently economic, and hence political forces affect a TA only *indirectly* through economic variables (e.g. outputs and trade volumes).

The third insight is that the presence of a domestic-commitment motive can explain why trade liberalization typically occurs in a gradual manner. In particular, the reduction in tariffs happens in two phases: first, there is an instantaneous drop in tariffs, which reflects the TOT motive for the TA, and subsequently there is a gradual tariff reduction, which reflects the domestic-commitment motive. Intuitively, the allocation distortions caused by protection are more severe in the long run than in the short run, and hence the domestic-commitment motive calls for bigger tariff reductions in the long run than in the short run. Furthermore, the speed of liberalization is increasing in z . The reason is that, if z is lower, the expected length of time for which capital owners are “stuck” in a sector is longer, so the lobby will insist on keeping a high protection level for a longer period of time.

Finally, [Maggi and Rodriguez-Clare \(2007\)](#) show that tariff ceilings are preferred to exact tariff commitments. The intuition is in two steps. First, if one focuses on complete TAs, the optimal exact tariff commitments in general are positive, though lower than the noncooperative levels, and hence induce allocation distortions. Second, consider replacing an optimal exact tariff commitment with a tariff ceiling at the same height: the former shuts down ex-post lobbying and contributions, while the latter leaves some discretion (governments have the *option* of setting tariffs below the ceilings) and hence induces ex-post lobbying and contributions; the latter is preferable because the anticipation of ex-post contributions reduces the expected net returns to capital in organized sectors, and hence mitigates the investment distortion. I will come back to the topic of tariff ceilings

and the incompleteness of TAs in Section 3, where I focus on the design of TAs, but here I note that Maggi and Rodriguez-Clare's model can explain why TAs are incomplete contracts without relying on the presence of contracting frictions between governments (although, as I highlighted above, contracting frictions between a government and its domestic actors are key).

Next I briefly discuss other papers that have highlighted domestic-commitment motives for TAs in the presence of lobbying. Mitra (2002) shows that a similar domestic-commitment motive as in Maggi and Rodriguez-Clare (1998) arises also in a setting where there is no long-run distortion in the capital allocation, but there is a resource cost of lobby formation: in this case, if the government does not commit to free trade, the long-run inefficiency generated by the prospect of trade protection (that the government does not get compensated for) is given by the cost of lobby formation. More broadly, Mitra's paper suggests that there may be a domestic-commitment motive for a TA any time the prospect of trade protection leads to a long-run misallocation of resources, whether it is in the form of misallocation of resources between productive activities or waste of resources in unproductive activities.¹⁵ Brou and Ruta (2009) extend Maggi and Rodriguez-Clare's (1998) model by allowing governments to use trade policies and domestic subsidies, and argue that the domestic-commitment theory of TAs can provide a rationale for the WTO's restrictions on the use of production subsidies.

Limão and Tovar (2011) propose a different version of the domestic-commitment argument for TAs. They consider a setting in which a small-country government bargains with a domestic lobby over two policy instruments, a tariff and a non-tariff barrier, where the latter is the less efficient redistributive instrument. In this setting they show that the government may benefit from committing to a tariff reduction because this may improve its bargaining position, and this benefit may outweigh the cost of constraining the more efficient redistributive tool. Finally, Liu and Ornelas (2012) argue that a TA can serve as a commitment device for the purpose of stabilizing a democratic regime. The key idea of this paper is that an incumbent government may value a TA because it leads to the destruction of rents, which in turn reduces the likelihood of a coup by rent-seeking autocratic groups, thereby helping consolidate unstable democracies.¹⁶

I conclude this section by mentioning another model where a government's lack of commitment vis-à-vis domestic agents has important implications for TAs. McLaren (1997) considers a two-period Ricardian model where a small country (S) negotiates a TA with a large country (L). In the first period, domestic agents commit their resources

¹⁵ Krishna and Mitra (2005) explore an interesting consequence of Mitra's argument: if a country liberalizes unilaterally because of a commitment issue as in Mitra (2002), this will have effects on the trade policies chosen by its trading partners. They show that it may induce a trading partner to reduce its own tariffs, because it increases the incentives for the export lobby in the partner country to form and lobby against the import-competing lobby there for lower protection.

¹⁶ This paper focuses on the rationale for joining free trade agreements, but the basic argument applies equally well to the case of a multilateral TA.

to a sector; in the second period, the governments negotiate over a tariff and a transfer through Nash bargaining. Given the resource allocation, the equilibrium TA involves free trade and a transfer from S to L. Ex-ante, anticipating free trade, agents commit resources to the sector where S has a comparative advantage. But this leads L to choose a higher tariff in the Nash equilibrium, which in turn worsens the outside option of country S in the trade negotiation. McLaren shows that this adverse effect of the *anticipation* of a TA on the welfare of the small country may outweigh the standard gains from trade, so this country may be better off by committing ex-ante *not* to sign a TA.

McLaren's point relates in an interesting way to the domestic-commitment theory of TAs. In McLaren's model, the TA can be interpreted as a short-term contract, because it occurs after investment decisions are made. But if the TA were a long-term contract, in the sense of occurring before investment decisions are made, then the hold-up problem highlighted by McLaren would not arise. Thus McLaren's model suggests that TAs can help *only if* they are effective long-run commitments (consistently with the domestic-commitment theory), while they can have perverse effects if they are only short-term commitments.¹⁷

2.3. New Trade Theories of Trade Agreements

A new and important line of research has emerged recently that explores the implications of imperfect competition for TAs. A central theme in this new area of research is that, in the presence of imperfect competition, TOT externalities are not the only international externalities from trade policy. In particular, three new externalities have been identified and examined: (i) "firm-delocation" externalities in the presence of free entry (Venables, 1985, 1987; Ossa, 2011; Bagwell and Staiger, 2009, 2012b), (ii) "profit-shifting" externalities (Mrazova, 2011; Ossa, 2012; Bagwell and Staiger, 2012a), and (iii) trade-volume externalities when prices are determined by bilateral bargaining (Antras and Staiger, 2012a,b). As I discuss below, these non-TOT externalities *may* be a separate cause of inefficiency in noncooperative policies, hence giving rise to new rationales for TAs, and can have important implications for the design of TAs.

2.3.1. Firm-Delocation and Profit-Shifting Externalities

In this section I focus on the implications of firm-delocation and profit-shifting externalities from trade policy, starting with the former type.

Venables (1985, 1987) was the first to identify the possibility of firm-delocation externalities from trade policies. This type of externality can arise whenever markets are imperfectly competitive and there is free entry. The basic idea is the following: if a

¹⁷ In a recent paper, Sovey (2012) develops a model where TAs are motivated by "political hold-up" problems. In her model, if a government makes a public investment in its comparative-advantage sector and hence makes itself more "dependent" on trade, it gives its trading partner an increased ability to extract political concessions in the future. As in McLaren (1997), a political hold-up problem calls for a long-term TA. Sovey then argues that a long-term TA is harder to self-enforce than a short-term one, because of the additional political uncertainty over the longer time horizon, and for this reason a multilateral institution like the WTO, by increasing the severity of punishments, can facilitate self-enforcement.

country imposes a tariff on imports, this will tilt the balance of competition in favor of domestic firms, and this in turn will induce exit of foreign firms and entry of domestic firms. In the presence of transport costs, this effect tends to benefit the country imposing the tariff and hurt the exporting country.

Ossa (2011) has explored the implications of firm-delocation externalities for the purpose and design of TAs. In particular, Ossa considers a Krugman-type model with monopolistic competition, CES preferences over varieties and iceberg transport costs. Governments maximize welfare and can only choose *ad-valorem* import tariffs. In this model, import tariffs have no TOT effects at all. Intuitively, firms apply a constant mark-up over marginal cost, so the incidence of an *ad-valorem* tariff falls entirely on the importing country, and hence ex-factory prices are unaffected.¹⁸ The feature that import tariffs have no TOT effects of course depends on the special model structure, but it serves to isolate the delocation externality, which operates in the following way: an increase in the Home tariff leaves the total number of domestic and foreign firms unchanged, but modifies its composition in favor of domestic firms; because of transport costs, this lowers the Home price index and increases the Foreign price index, thus leading to an increase in Home welfare and a decrease in Foreign welfare.¹⁹

As a consequence of the negative delocation externality that a tariff exerts on the exporting country, the noncooperative equilibrium entails inefficiently high tariffs, and so there is scope for a TA to reduce tariff levels. Ossa argues that this rationale for TAs resonates with the often-heard informal argument that import protection leads to a loss of manufacturing firms and “good jobs” in the exporting country, and the role of a TA is to prevent governments from engaging in this beggar-thy-neighbor behavior.²⁰

Mrazova (2011) and Ossa (2012) focus on the implications of a different type of policy externality that may arise under imperfect competition, namely the *profit-shifting* externality. The profit-shifting effect of trade policies was first studied by Brander and Spencer (1984, 1985) in the context of a Cournot oligopoly with a fixed number of firms. Mrazova (2011) focuses on a setting similar to Brander and Spencer’s, while Ossa (2012) focuses on a monopolistic competition model with a fixed number of firms. In

¹⁸ Terms of trade in this setting can be defined in two different ways: as the ratio between the ex-factory price of a foreign variety and that of a domestic variety, or as the ratio between the price index for exported varieties and that for imported varieties. Ossa shows that with the first definition tariffs do not affect TOT, and with the second definition a tariff *worsens* the country’s TOT. In the text I am implicitly adopting the first of these two definitions. It is also important to point out that, while import tariffs have no TOT effects (according to the first definition above), export taxes would have dollar-for-dollar effects on TOT, as emphasized by Bagwell and Staiger (2009).

¹⁹ There is also a counteracting effect, because the tariff makes foreign products more expensive for consumers, but this effect is shown to be dominated.

²⁰ Ossa also argues that his model can provide a rationale for GATT’s rules of reciprocity and MFN, much in the same way as a TOT model; I will come back to this aspect in Section 3.

both models, governments can only use import tariffs.²¹ Just like the TOT externality, the profit-shifting externality from a tariff is negative (holding TOT fixed, an increase in the tariff hurts the exporting country), and as a consequence, a mutually beneficial TA must reduce tariff levels relative to the noncooperative equilibrium. Mrazova (2011) in addition shows that the Bagwell-Staiger “test” fails in her setting (PO tariffs are inefficient), thus the purpose of a TA indeed goes beyond the correction of TOT externalities.

Enter Bagwell and Staiger. In two companion papers (Bagwell and Staiger, 2009, 2012a) they argue that, if import instruments *and* export instruments are available, even in the presence of delocation or profit-shifting externalities the only purpose of a TA remains the correction of TOT externalities. They consider a number of possible market structures, including monopoly, monopolistic competition, and Cournot oligopoly (with or without free entry, and with or without integrated markets), and show that, if countries can use both import and export taxes and there are no income effects (quasi-linear preferences), then PO policies are efficient. On this basis, Bagwell and Staiger conclude that neither delocation nor profit-shifting externalities constitute a “fundamental” rationale for TAs.²²

I will next try to illuminate the logic of Bagwell and Staiger’s argument by considering a slightly more general setting. Focus on a two-country world with any number of goods, and suppose governments can choose specific trade taxes.²³ With a slight abuse of notation, $(\tau, \tau^*, p, p^*, p^w)$ will now denote the *vectors* of trade taxes and prices.

Government objectives can always be expressed in reduced form as functions of trade taxes ($\Omega(\tau, \tau^*)$ and $\Omega^*(\tau, \tau^*)$), and trade taxes in turn can be written as price wedges ($\tau = p - p^w$ and $\tau^* = p^* - p^w$), so government objectives can always be expressed as functions of local, foreign, and world prices ($\tilde{\Omega}(p, p^*, p^w) \equiv \Omega(p - p^w, p^* - p^w)$ and $\tilde{\Omega}^*(p, p^*, p^w) \equiv \Omega^*(p - p^w, p^* - p^w)$), regardless of the nature of the international policy externalities. Importantly, note that this setting allows for virtually any underlying market structure.²⁴ Note also that a government objective may depend on local prices in both

²¹ Mrazova justifies the assumption that only tariffs are available by proposing a complementary theory that explains why export subsidies have been banned by the GATT-WTO. I will be more specific on this part of her theory below, where I focus on possible resolutions of the export subsidy puzzle.

²² Ossa has replied to this criticism by observing that in reality the use of export instruments is severely restricted: (i) export subsidies have been banned by GATT a long time ago, and the subsequent rounds of negotiations have focused mostly on import barriers, and (ii) the US has banned export taxes by constitution. My personal opinion is that Bagwell and Staiger are correct in pointing out that a complete theory should in principle explain, not assume, the pre-existing restrictions on export instruments. However, it is not hard to imagine a model where there are transaction costs or political frictions such that trade negotiations do not address import barriers and export instruments simultaneously in a single round but rather in a sequential manner, or such that the unilateral use of export instruments is subject to frictions. In such a richer model, delocation (or profit-shifting) externalities would indeed be a distinct motive for TAs.

²³ The argument can be easily extended to allow for production and consumption taxes.

²⁴ Note that I have implicitly made two assumptions for simplicity. The first is that a government can apply different trade taxes for different goods. If goods are differentiated and a government must apply the same trade tax on all the

countries, and this is the new feature relative to the perfect-competition setting described in Section 2.1. This is a key point to keep in mind as we proceed: *non-TOT externalities can always be seen as local-price externalities.*

Assume that there are *no income effects*. Together with the assumption that there are only two countries, this implies that the local price of each good depends only on the *total* trade tax on that good. This feature is the key to Bagwell and Staiger's argument: under the assumptions I just stated, the import tax and the export tax on a given good are perfectly substitutable in affecting local prices.²⁵

The PO policies are defined as the ones that would result if governments did not value changes in p^w . Since local prices depend only on total trade taxes (denoted $\bar{\tau} \equiv \tau + \tau^*$), this means that at the political optimum governments are effectively choosing the same variables, $\bar{\tau}$. Since the common choice of $\bar{\tau}$ must maximize both the Home objective and the Foreign objective, and since changes in p^w act as pure transfers, it follows that PO policies are efficient. Formally, PO policies are defined by the first order conditions $\tilde{\Omega}_p p_{\bar{\tau}} + \tilde{\Omega}_{p^*} p_{\bar{\tau}}^* = 0$ and $\tilde{\Omega}_p^* p_{\bar{\tau}} + \tilde{\Omega}_{p^*}^* p_{\bar{\tau}}^* = 0$ (where the notation has the intuitive meaning). Defining global welfare as $\tilde{\Omega} + \tilde{\Omega}^*$, and noting that a change in p^w does not affect global welfare (it is a pure transfer), the FOC for global efficiency is $(\tilde{\Omega}_p + \tilde{\Omega}_p^*) p_{\bar{\tau}} + (\tilde{\Omega}_{p^*} + \tilde{\Omega}_{p^*}^*) p_{\bar{\tau}}^* = 0$. Clearly, the PO policies satisfy the FOC for global efficiency.

I summarize this discussion with the following:

Efficiency of Political Optimum (EPO). Assume: (i) there are only two countries; (ii) there are no income effects; and (iii) governments choose only trade taxes. Then PO policies are efficient, even in the presence of local-price externalities.

This result is in a way very general and in a way very special. The sense in which it is very general is that it holds regardless of the nature of international policy externalities. Indeed, it holds even in scenarios where intuition might suggest that there are other motives for a TA beyond the correction of TOT externalities. For example, suppose there are non-pecuniary international externalities, for example because of cross-border pollution: the argument above is still valid, and hence PO policies are efficient. To be more concrete, consider a perfectly-competitive, partial-equilibrium model with a single good, where governments maximize welfare and there is a cross-border pollution externality generated by production in the Foreign country. Let $\Upsilon(x^*)$ denote the environmental damage caused in the Home country by foreign production and $x^*(p^*)$

varieties of that good, the argument must be slightly adapted, but it still goes through. The second assumption is that markets are integrated, so arbitrage implies that there is a single world (offshore) price for each good. If markets are segmented there may be two-way trade in identical commodities, and so there may be two distinct offshore prices for the same good, one for each direction of trade; but again, the argument is easily extended to cover this case.

²⁵ Intuitively, note first that the wedge between local prices is given by the total trade tax ($p - p^* = \tau + \tau^*$). Next note that changing τ and τ^* in a way that leaves the total trade tax constant causes a transfer of revenue between the countries. If there are no income effects, this transfer of revenue will not affect demand functions, thus only the total trade tax matters for equilibrium local prices, not its composition.

the foreign supply function. If Home welfare net of environmental damage is given by $W(p, p^w) - \Upsilon(x^*(p^*)) \equiv \Omega(p, p^*, p^w)$, and Foreign welfare is defined analogously, then the EPO result above immediately applies.²⁶

At the same time, the three assumptions (i)–(iii) are very restrictive, and even though the EPO result only states a sufficient condition, each of the three assumptions plays a key role in delivering the efficiency of the PO policies: when local-price externalities are present, if any of the three assumptions is violated then import-side policies and export-side policies in general are not perfectly substitutable in affecting local prices, and hence PO policies will typically be inefficient.²⁷

Before concluding this subsection, I return one last time to the “export subsidy puzzle.” In Section 2.2, I discussed how the domestic-commitment theory can provide a possible resolution to this puzzle. As I discuss next, also models with firm-delocation and profit-shifting externalities offer avenues to address the puzzle.

Bagwell and Staiger (2012b) consider a linear Cournot delocation model à la Venables (1985), where two governments choose trade taxes to maximize welfare. First they show that, starting from free trade, a country acting unilaterally can increase its welfare with a small export subsidy (because its beneficial delocation effect outweighs its adverse TOT effect) and with a small import tariff. This suggests that imposing a cap on export subsidies may be jointly beneficial for the two countries. However, it turns out that the Nash equilibrium involves both import taxes and export taxes; what is responsible for this surprising result is the fact that import and export taxes are complementary (increasing the tariff makes an export tax more attractive). Thus, if governments negotiate over import and export instruments starting from the Nash equilibrium, the model is not able to explain why a TA would cap export subsidies. But if negotiations initially focus only on import tariffs, bringing them close enough to zero, in a subsequent phase of negotiation there will be scope for imposing a cap on export subsidies. In a similar vein, DeRemer (2011) considers a monopolistic competition model where governments can choose trade taxes and production subsidies. He argues that capping both export subsidies

²⁶ The EPO result applies also if, in addition to trade taxes, governments can use production taxes, which are the first-best instrument to correct the environmental externality (see footnote 23).

²⁷ To be clear, there may be special circumstances in which PO policies are efficient even if some of the conditions (i)–(iii) are not satisfied. For example, Bagwell and Staiger (2012a) show that PO policies are efficient in a special three-country setting with competing exporters. And in Antras and Staiger (2012a), as I will mention later, PO policies are efficient in the special case where governments maximize welfare, in spite of there being three countries in their model. But these cases are rather special, and that is why I use the word “typically” in the text. Here I should also highlight the relationship between my EPO result and a result shown by Bagwell and Staiger (2012a): they show that, if conditions (i)–(iii) above are satisfied *and* (iv) government preferences can be represented as functions of world and local prices, then PO policies are efficient. Importantly, they emphasize that condition (iv) can only be assessed given the specific economic structure, and they check that this is the case in a number of imperfect-competition models. The value added of my EPO result is to show that, if conditions (i)–(iii) are satisfied, there is no need to know anything else about the economic structure or the nature of the international externalities to conclude that PO policies are efficient.

and production subsidies may be desirable, but only if import tariffs are sufficiently close to zero, not if they are close to their Nash equilibrium levels.

While the two models mentioned above are broadly suggestive of reasons why the GATT-WTO has banned export subsidies, neither of them can explain an outcome where export subsidies are present in the noncooperative equilibrium *and* a TA bans export subsidies. [Mrazova's \(2011\)](#) model can explain both of these features. Her model allows for political economy considerations, so the Nash equilibrium may entail export subsidies. The basic idea to explain the ban on export subsidies is the following. Recall from Section 2.1.1 that an efficient TA determines only net trade taxes, not the separate import and export tax levels (see equation (3)) so that a given point on the efficiency frontier can be achieved with import instruments alone or with export instruments alone. Assuming a fixed cost of administering each policy instrument, efficiency requires the use of (at most) one policy instrument for each good. Mrazova then considers a repeated-game model of TAs and argues that, due to profit-shifting effects, a tariff-only agreement is more easily self-enforced than a subsidy-only agreement, so an export subsidy ban is desirable.

2.3.2. Trade-Volume Externalities Due to Bilateral Bargaining

When prices of international transactions are determined by bilateral bargaining rather than by market clearing, the international externalities exerted by trade policy are of a different nature than the ones highlighted so far, and have novel implications for the purpose and design of TAs. This is the focus of two recent papers by [Antras and Staiger \(2012a,b\)](#).

It is convenient to start with [Antras and Staiger \(2012b\)](#), which focuses on a simple matching model to highlight some key implications of bilateral bargaining for TAs. More specifically, this paper considers a two-country, partial-equilibrium model where all producers are located in the Foreign country. Each producer is matched with a consumer, and within each match the quantity exchanged and the price are determined by bilateral bargaining. Governments maximize welfare, with the Foreign government choosing an export tax and a domestic input subsidy, and the Home government choosing an import tariff.

In this environment, international policy externalities cannot be viewed as traveling simply through TOT. To understand the key difference between this environment and a setting with market clearing, notice first that in the case of market clearing, Foreign policies can affect the point of Home's offer curve that is selected in equilibrium, but cannot affect Home's offer curve itself; so they can affect world price and trade volume but cannot control them separately; in this sense, there is a single channel of international policy externality, which can be viewed alternatively as a TOT externality or a trade-volume externality. With bilateral bargaining, on the other hand, this link is broken, and Foreign policies can affect TOT and trade volume separately, thus the rationale for a TA is to jointly correct these two separate externalities. Indeed, with bilateral bargaining, PO

policies can be shown to be inefficient, so the purpose of a TA goes beyond the correction of TOT externalities.²⁸

Pricing by bilateral bargaining is particularly relevant when firms offshore the production of specialized inputs and there is incomplete contracting between downstream and upstream producers. The implications of offshoring for TAs are the focus of [Antras and Staiger \(2012a\)](#). This paper considers an environment with three countries: Home, Foreign, and the rest of the world (ROW). Home is the sole producer of a final good that requires a custom-made intermediate input; Foreign is the sole producer of the intermediate input; and ROW specializes in a plain-vanilla numeraire good. Once a Foreign upstream firm and a Home downstream firm are matched, the former must customize the intermediate input for the latter, and then the price of the input is determined by bilateral bargaining. The customization of the input cannot be contracted upon ex-ante, thus a standard hold-up problem arises. Each government chooses trade taxes to maximize a possibly politically adjusted social welfare function. In this setting, beyond standard TOT externalities, trade policies exert trade-volume externalities of the kind described above, with the additional feature that, by affecting trade volumes, a country's trade policies can affect the severity of the hold-up problem in the other country.

In this environment, [Antras and Staiger \(2012a\)](#) show that PO policies are in general inefficient, except in the special case where governments maximize welfare. Thus, as long as there are political-economy considerations in the governments' objectives, the rationale for TAs goes beyond the correction of TOT externalities.²⁹

It is clear that, in this setting, the international externalities from trade policies do not simply travel through TOT. However, as the EPO result of Section 2.3.1 highlights, the presence of non-TOT externalities is not sufficient to conclude that PO policies are inefficient. So why does the EPO result *not* apply here? The key reason is that in this setting there are more than two countries, and as a consequence, export taxes and import taxes are not perfectly substitutable in affecting local prices.³⁰ To confirm this point, I note that if [Antras and Staiger's \(2012a\)](#) model were played out in a two-country world, then the PO policies would be always efficient, regardless of the government objectives. Thus in

²⁸ The implication that PO policies are inefficient is not highlighted in the paper, but it is easy to show. The paper focuses on a design question, namely whether a "shallow integration" approach can achieve global efficiency, but the paper is relevant also for the question of the purpose of a TA. I will come back to the design question in Section 3. Also, it is interesting to note that the EPO result presented in Section 2.3.1 does *not* apply in [Antras and Staiger's \(2012b\)](#) setting, because trade taxes are not the only policy instruments: Foreign can use also a non-wedge policy, namely an input subsidy, and there is no Home policy that is a perfect substitute for it in affecting local prices.

²⁹ A second question examined by [Antras and Staiger \(2012a\)](#) concerns the desirability of a "shallow integration" approach. Again, I will focus on this design question in Section 3.

³⁰ Intuitively, recall that Home is the sole producer of the final good, so the local price of this good in the Home market (say p^H) depends on all three trade taxes on this good, but not through their sum, because Home's export tax has a larger impact on p^H than each of the other countries' import tax. Also note that the reason why the EPO result does not apply in [Antras and Staiger \(2012a\)](#) is different than the one in [Antras and Staiger \(2012b\)](#). As highlighted in footnote 28, in the latter model (which has only two countries) the reason the EPO result does not apply is that trade taxes are not the only policy instruments.

some sense the more surprising aspect of the results in [Antras and Staiger \(2012a\)](#) is not that the PO policies are inefficient in the presence of political-economy considerations, but rather that the PO policies *are* efficient in the case of welfare-maximizing governments.

2.4. The Uncertainty-Managing Motive for a TA

The papers discussed so far abstract from the presence of uncertainty, and highlight the gains that a TA can offer by changing the levels of trade policies relative to the non-cooperative equilibrium. But if the political/economic environment is uncertain, one can distinguish between an “uncertainty managing” motive and a “mean managing” motive for a TA. Suppose that, because of shocks in the political/economic environment, noncooperative trade policies are themselves subject to shocks. We can then ask the following question: can governments achieve mutual gains by changing the degree of uncertainty in trade policies relative to the noncooperative equilibrium, *holding their mean levels constant*? If the answer is yes, we can say that there is an “uncertainty managing” motive for a TA. [Limão and Maggi \(2013\)](#) examine under what conditions there exists an uncertainty-managing motive for a TA, whether it calls for a reduction or an increase in policy uncertainty, and what are the potential gains from a TA that regulates policy uncertainty.³¹

In a standard competitive trade model with risk neutrality, where trade policies exert international externalities only through TOT, Limão and Maggi show that there tends to be an uncertainty-*increasing* motive for a TA, due to the convexity of indirect utility and revenue functions in prices. This model thus seems at odds with the often-heard argument that TAs can provide welfare gains by reducing trade policy uncertainty. When individuals are risk averse, on the other hand, the direction of the uncertainty motive for a TA is determined by a trade-off between risk aversion and flexibility: the degree of risk aversion, in interaction with the degree of openness, favors an uncertainty-reducing motive; while the degree of flexibility of the economy, which is in turn determined by the export supply elasticity and the degree of production diversification, favors an uncertainty-increasing motive. Empirically lower-income countries tend to have lower export supply elasticities and a lower degree of diversification, thus Limão and Maggi’s model suggests that the uncertainty-reducing motive might be relatively more important for lower-income countries. Another key result is that, as trade costs decline, the gains from reducing trade policy uncertainty tend to become more important relative to the gains from reducing average trade barriers. A broad implication of this finding is that

³¹ Policy makers and practitioners often argue that one of the main goals of TAs is to reduce uncertainty in trade policies, and various TAs including the WTO include “mission” statements that can be interpreted along similar lines. For example, the WTO states in its website that “Just as important as freer trade—perhaps more important—are other principles of the WTO system. For example: non-discrimination, and making sure the conditions for trade are stable, predictable and transparent.”

uncertainty-reducing motives for TAs are likely to emerge as the world becomes more integrated, and are more likely to be present for countries within a region.³²

2.5. Empirical Evidence

The empirical literature on TAs is still in its infancy, but it has seen a considerable acceleration in the last decade or so. In this section I discuss some papers that attempt to get at the underlying motives for TAs, and some that examine the impacts of TAs on trade barriers and trade flows in a more descriptive way. I postpone a discussion of the empirical work on regional trade agreements to Section 4.

2.5.1. Tests of the TOT Theory

A number of recent papers have set out to test the predictions of the TOT theory. Four papers stand out in this group. The first one is [Broda et al. \(2008\)](#), who focus on the prediction that, in a noncooperative scenario, tariffs should tend to be higher for countries/goods where market power (the inverse of the export supply elasticity) is higher. Broda et al. consider the tariffs set by 15 non-WTO countries, on the presumption that these countries choose trade policies in a noncooperative manner. They estimate export supply elasticities by country and good—a significant contribution in itself—and find that these elasticities are related with tariffs in the way predicted by the theory, particularly if one focuses on the variation across goods within a country. Next they control for political-economy determinants of tariffs, using a parsimonious specification à la Grossman–Helpman (with the additional assumption that all industries are politically organized), and find that export supply elasticities retain explanatory power even in the extended specification.³³

[Bagwell and Staiger \(2011\)](#) test the predictions of the TOT model regarding the tariff cuts that a country should make when *acceding* the WTO. Bagwell and Staiger start by showing that, if demand and supply functions are linear, the model predicts that the tariff cut should be deeper, other things equal, when the noncooperative volume of imports is higher. They then test this prediction across six-digit HS level goods and across 16 countries that acceded the WTO between 1995 and 2005, finding a strong positive correlation. The correlation survives also in the presence of country and good

³² Also [Handley \(2012\)](#) and [Handley and Limão \(2012, 2013\)](#) focus on the links between uncertainty and TAs, but from a very different perspective. They examine (theoretically and empirically) the impact that TAs have on trade flows by removing the risk of future increases in protection, taking trade policy (before and after the TA) as exogenous. Handley and Limão do not consider the key questions addressed in [Limão and Maggi \(2013\)](#), namely whether there is “too much” or “too little” uncertainty in the noncooperative trade policies and what are the gains from “correcting” the degree of policy uncertainty through a TA. I will come back to the papers by Handley and Limão in the next section.

³³ However, there is one finding in [Broda et al. \(2008\)](#) that is not easy to reconcile with the TOT theory. According to the theory, a country acting noncooperatively should set discriminatory tariffs, because export supply elasticities vary across exporters, but this almost never happens in the data. The authors argue that the presence of administrative costs can reconcile this observation with the theory, but these costs would have to be very high, because export supply elasticities vary widely across exporters and therefore the potential gains from discrimination are high.

fixed effects, and importantly, it passes the “placebo” test that it should hold only for tariffs on imports from other WTO members, not on imports from non-WTO countries.³⁴

The papers discussed above focus on non-WTO countries (Broda et al.) or countries that recently joined the WTO (Bagwell and Staiger), so they leave out the vast majority of current WTO countries. Ludema and Mayda (2010) focus instead on the MFN tariffs of all WTO members. Their test of the TOT theory is based on the following idea: the MFN rule causes a well-known free-rider problem in multilateral negotiations, and for this reason negotiations are only partially successful in removing TOT considerations from tariff levels, therefore the negotiated MFN tariffs should still partially reflect the market power of importing countries. Moreover, the correlation between MFN tariffs and market power should be stronger when exporter concentration (as measured for example by the Herfindahl index) is lower, because in this case the free-rider problem is more severe, thus MFN tariffs should be negatively related to the product of exporter concentration and importer market power. Ludema and Mayda test this prediction on a cross-section of MFN tariffs set by WTO members in the Uruguay Round, finding supportive results.

Finally, Bown and Crowley (2013a) test the predictions of a repeated-game version of the TOT model (namely, Bagwell and Staiger’s (1990) model of “managed trade”) using data on US temporary tariffs imposed under the US’s antidumping and safeguard laws over 1997–2006. The key idea of the model is that, if a TA is to be self-enforcing, it needs to provide for “escape clauses” that allow countries to raise tariffs in periods when the temptation to defect from the agreement is stronger, that is when the incentive to manipulate TOT is stronger, which in turn tends to happen when trade volumes are higher.³⁵ Thus a key prediction of the model is that temporary tariffs should be observed with higher likelihood when import volumes are higher. Bown and Crowley find strong support for this prediction in the data.

Finally, I should mention that there are a number of empirical studies documenting that a country’s tariffs can significantly affect its TOT, which of course is a pre-requisite for the empirical relevance of the TOT theory. Papers in this group include Kreinin (1961), Winters and Chang (2000, 2002), and Bown and Crowley (2006).

2.5.2. Tests of the Domestic-Commitment Theory

As a whole, the studies discussed above are quite supportive of the TOT theory. At the same time, I do not think this body of research has established that the TOT motive is the *only* empirically significant motive for TAs. This leads me to the next question, which is whether other motives for TAs are empirically important. The short answer to this

³⁴ Bagwell and Staiger also consider a more general relationship predicted by TOT theory between tariff cuts, trade elasticities and import volumes, and test this prediction using Broda et al.’s (2008) trade elasticities for the five countries within their sample for which such elasticities are available. Also the results of this test are consistent with the predictions of the theory.

³⁵ See Section 3.2.1 for further discussion of repeated-game versions of the TOT model.

question is that we do not know yet: domestic-commitment theories and New Trade theories of TAs have thus far received less empirical attention than the TOT theory, and the jury is still out. I will start by focusing on the empirical research on the domestic-commitment theory.

The first paper in this area is by [Staiger and Tabellini \(1999\)](#), who test a prediction of their theoretical model ([Staiger and Tabellini, 1987](#)), in which the government is subject to a time-inconsistency problem due to the fact that it chooses trade policy after domestic agents have made their allocation decisions. This model suggests that, if the government commits to a TA to address this time-inconsistency problem, the TA should lead to deeper trade liberalization in sectors where the potential for production distortions from protection is larger. Staiger and Tabellini test this prediction by focusing on the sectoral exclusions chosen by the US government in the Tokyo Round of GATT, using as “control” group the US tariff decisions made under the GATT’s escape clause, which arguably were not effectively constrained by GATT. Their findings are broadly supportive of the model’s prediction.

[Limão and Tovar \(2011\)](#) test their theoretical model (see Section 2.2) using data on tariffs and non-tariff barriers (NTBs) in Turkey. One key prediction of their model is that a government is more likely to commit to a tariff cap in industries where its bargaining power relative to the lobby is lower, and conditional on committing, the tariff cap should be tighter when the government’s bargaining power is lower. A key ingredient in testing this prediction is measuring the government’s relative bargaining power industry by industry. To do so, Limão and Tovar posit that the relative bargaining power of the government in a given industry is lower when the rate of firm exit in that industry is lower. The idea is that, if the exit rate is lower, the firms (and the lobby that represents them) discount the future less, while the government’s discount rate does not vary across industries, and noncooperative bargaining theory suggests that a player’s relative bargaining power is higher when his or her relative patience is higher. Using their estimates of relative bargaining powers, Limão and Tovar find that in the Uruguay Round the Turkish government indeed committed to less stringent tariff bindings in industries where its relative bargaining power was stronger, and did not commit at all if the latter was strong enough.³⁶ It is interesting to note that this finding is broadly consistent also with the predictions of [Maggi and Rodriguez-Clare’s \(1998, 2007\)](#) version of the domestic-commitment theory.³⁷

[Liu and Ornelas \(2012\)](#) test their theory that a TA may serve as a commitment device for a fragile democracy to destroy protectionistic rents and hence reduce the likelihood of

³⁶ Another finding in [Limão and Tovar \(2011\)](#) is that Turkey used NTBs more heavily after the tariff bindings were imposed. This finding as well is consistent with their model, but I note that this kind of policy substitution between tariffs and NTBs is consistent also with a variety of models that do not feature domestic-commitment motives for a TA.

³⁷ In [Maggi and Rodriguez-Clare \(1998\)](#) the government is more likely to commit to free trade when its bargaining power is lower, and in [Maggi and Rodriguez-Clare \(2007\)](#) tariff caps tend to be tighter when governments have lower bargaining power (see section I.D of that paper).

coups by rent-seeking authoritarian groups (see Section 2.2), by using data on preferential trade agreements for 116 countries over the period 1960–2007. In line with their model’s predictions, Liu and Ornelas find that more fragile democracies are indeed more likely to sign preferential TAs, and that signing a preferential TA in turn lowers the likelihood of democratic failure.

I would summarize the thin empirical literature on the domestic–commitment theory of TAs by saying that it has found support for *some* predictions of *some* versions of the theory, but a broad and systematic empirical investigation of this theory is still missing. Ultimately, the hope is to be able to quantify the relative importance of TOT motives and domestic–commitment motives for TAs, but this is certainly no easy task.

2.5.3. *Empirical Work on the New Trade Theory*

Empirical research focused on New Trade theories of TAs is at the very beginning. I am not aware of any attempts to test these theories with econometric approaches, but a recent paper by Ossa (2013) takes the theory to the data using a calibration approach.

Ossa develops a multi-country model that allows for three drivers of trade protection: TOT effects, profit-shifting effects, and political-economy considerations.³⁸ The model, which combines elements from Krugman (1980) and Grossman and Helpman (1995a), is calibrated to match observed trade flows and tariffs at the industry level in 2005. Using a technique introduced by Dekle et al. (2007), Ossa performs counterfactual analysis using only estimates of the elasticities of substitution (taken from Broda and Weinstein, 2006), estimates of political-economy weights (taken from Goldberg and Maggi, 1999), and factual levels of trade flows and tariffs. Several interesting findings arise. First, TOT and profit-shifting drivers of protection quantitatively dominate political-economy drivers. Second, a global trade war would lead to tariffs averaging about 60% across industries and countries, and would reduce welfare by about 3.5% relative to the cooperative outcome. Finally, relative to where we are today, the potential gains from further multilateral trade negotiations are negligible.

Whether these are “numbers we can believe in” is not obvious, given the very stylized nature of the model, but this is a thought-provoking paper that points to a promising way forward for addressing important questions such as quantifying the relative importance of different motives for trade protection, the gains achieved by past TAs, and the potential gains from future TAs.

2.5.4. *Impacts of the GATT-WTO*

In this subsection I briefly discuss a recent wave of papers that have examined the impact of the GATT-WTO on trade barriers and trade flows. This literature was triggered by

³⁸ An earlier attempt at quantifying New Trade motives for protection can be found in Ossa (2011), in the context of a model featuring only firm–delocation effects. In that calibration exercise, Ossa allows each country to set only a single tariff on all imports, and only focuses on noncooperative tariffs.

Rose (2004a), who sent shockwaves through the trade policy community (academic and not) by arguing that the WTO had virtually no impact on trade flows, based on a simple reduced-form regression analysis. In a similar vein, Rose (2004b) argued that the WTO had a negligible effect on the trade policies actually applied by countries.

These papers spawned a number of follow-up studies, most of which qualified Rose's results in significant ways. Subramanian and Wei (2007) show that the impact of WTO has been very uneven across countries and sectors, for example because developing countries enjoyed exemptions from trade liberalization in specific sectors (such as textiles); once these exceptions are accounted for, the WTO is found to significantly promote trade. Tomz et al. (2007) argue that many countries were mistakenly classified as non-members of the GATT, while in reality they were *de facto* members with similar rights and obligations as formal members, and show that this misclassification leads to underestimating the effect of GATT on trade flows. Liu (2009) shows the importance of “zeroes” in bilateral trade flows: if one takes into account that the WTO has led new country pairs to initiate bilateral trade—the “extensive partner-level margin” of trade—then the WTO is found to have a significant positive effect on trade. Dutt et al. (2011) find that the impact of WTO membership is significant on the extensive product margin of trade, that is, WTO membership leads to an increase in the number of goods traded, but the impact of the WTO is negligible on the intensive margin (the trade volume of already-traded goods).³⁹

Next I discuss some recent papers that also examine the effects of TAs on trade flows, but use more structural approaches, and provide some evidence about the mechanisms through which a TA affects trade.

Eicher and Henn (2011) examine the effects of WTO and regional trade agreements on trade flows by considering a panel of 177 countries over 50 years. They start with a reduced-form gravity approach that encompasses the specifications by Rose (2004a), Tomz et al. (2007), and Subramanian and Wei (2007), and find that only regional trade agreements have a significant impact on trade, not the WTO. Then they consider an augmented gravity equation that incorporates a key effect suggested by the TOT theory, namely that countries with more market power should agree to bigger tariff cuts as they join the WTO, and hence their trade volumes should increase by more. When a measure of market power (pre-accession import volumes) is incorporated in the regressions, the WTO is found to have a significant effect for countries with import volumes above the 85th percentile. This finding contributes to reconcile the seemingly contradictory results of reduced-form studies à la Rose (2004a) and theory-driven studies à la Bagwell and Staiger (2010a).

Finally, Handley (2012) and Handley and Limão (2012, 2013) show that the mechanisms through which TAs affect trade flows may be more subtle than just a decrease of tariff *levels*. These papers argue that, when trade policies are subject to shocks, export-

³⁹ Another paper worthy of mention is Tang and Wei (2010), which examines the impact of WTO on GDP growth, finding that WTO membership tends to be associated with higher GDP growth rates for developing countries.

ing firms respond not just to changes in the applied levels of trade barriers, but also to changes in the probability that trade barriers might be raised in the future. Thus, by reducing the risk of future protectionist spikes, a TA may encourage investment in export markets and boost trade volumes even if no change in applied policy levels is observed. [Handley \(2012\)](#) focuses on Australia's accession to WTO, finding evidence that this caused an increase in exports to Australia more because committing to WTO bindings removed the risk of future "bad news" for exporters, than because of actual reductions in Australia's applied tariffs. [Handley and Limão \(2012\)](#) find evidence that Portugal's accession to the EC boosted Portuguese exports to other EC countries in spite of the fact that Portugal already enjoyed free access to those countries before accession, thanks to pre-existing preferences, and estimate that a significant fraction of this effect was due to the fact that accession to the EC eliminated the risk faced by Portuguese exporters of losing pre-existing preferences. Finally, [Handley and Limão \(2013\)](#) estimate that a significant portion of China's rapid increase in exports to the US starting in 2001 is explained by the permanent MFN status gained by China as a consequence of its WTO accession, which removed the US threat of imposing "column 2" tariffs on imports from China.

3. THE DESIGN OF TRADE AGREEMENTS

In a world without transaction costs, the issue of how to design a TA would be rather uninteresting. Imagine for a moment that governments could costlessly write a complete contract that covers all relevant policies and contingencies and can be perfectly enforced. In such a world (which is the world I effectively considered in the previous section) there would be no need to think hard about how to design a TA: governments would be able to achieve a fully efficient outcome by writing a complete contract. Such contract would contain a large amount of detail, but would be conceptually straightforward.

In this section I will discuss the literature on the design of TAs as viewed from the perspective of transaction costs. To be a bit more specific, I will use the expression "transaction costs" as an umbrella term that encompasses two categories of frictions: (1) *contracting frictions*, which include costs of negotiating and writing contracts, imperfect verifiability, and private information, and (2) *imperfect enforcement*, by which I mean the lack of external enforcement power, so that TAs must be self-enforcing contracts.

Note that, while contracting frictions naturally lead to contract incompleteness, meaning that relevant contingencies and/or policies are missing from the contract, enforcement frictions typically do *not* generate contract incompleteness: the presence of self-enforcement constraints *per se* is not a reason for removing contingencies or policies from the contract. In fact, it can have the opposite effect, in the sense of inducing governments to introduce contingencies that otherwise would not be present in the contract. For example, self-enforcement constraints may require the agreement to be contingent on import shocks (escape clauses), whereas a perfectly enforceable agreement

would not need to be contingent.⁴⁰ For this reason, below I organize my discussion in two parts: first I will focus on the implications of contract incompleteness for TA design, and then I will focus on the implications of imperfect enforcement.

What would we miss if we ignored the presence of transaction costs? Why not stop at the world considered in the previous section, where the only questions concern the motives for a TA and how the efficient policies differ from noncooperative policies? If theory stopped there, it would not be able to explain a number of important features of real-world TAs. For example, it would be hard to explain why the GATT-WTO specifies tariff caps instead of exact tariff commitments; or why many domestic policies such as standards or domestic taxes are left to the governments' discretion; and so on.

Also, if we ignored transaction costs it would be difficult to explain the nature of trade disputes in the WTO and the role of the WTO court, the Dispute Settlement Body (DSB): absent transaction costs the only possible role for a court would be to enforce the obligations specified in the agreement, but in reality trade disputes are more often about the interpretation of vague obligations, or instances for which the agreement is silent, than about the enforcement of clear obligations. Thus a potentially important role of the DSB is to "complete" various dimensions of an incomplete contract, and therefore designing the role of the DSB becomes of key importance.

Before proceeding, it is useful to distinguish between three dimensions of TA design: (1) the design of substantive *policy rules*, that is constraints on governments' policy choices (e.g. tariff ceilings, non-discrimination rules); (2) the design of *enforcement rules* (how should governments behave after a violation of the agreement?); and (3) the design of *procedures*, such as dispute settlement procedures.⁴¹ I will argue that introducing transaction costs in our conceptual frame can take us a long way toward understanding the design of TAs along these three dimensions.

3.1. Contract Incompleteness and Trade Agreements

The overarching theme of this section is that viewing TAs as incomplete contracts can help understand the way TAs are designed. In the models I will discuss below, the incompleteness of the TA is sometimes derived endogenously from contracting frictions, sometimes assumed exogenously, and sometimes left implicit. I would argue that modeling contracting frictions *explicitly* has the advantage of forcing us to think in a disciplined way about the rationale for the rules and procedures we observe in real-world TAs. I am not advocating a dogmatic approach where everything must be explained from "first principles," but I think that too often a theoretical "story" proposed to explain a certain rule

⁴⁰ This is the case for example in Bagwell and Staiger's (1990) model of "managed trade" (see Section 3.2.1).

⁴¹ Another important type of procedure is the bargaining protocol for trade negotiations. In spite of the obvious importance of negotiation protocols, however, I am not aware of any formal literature addressing this question. On a distinct note, a paper that focuses on questions of procedure but does not fit easily in the taxonomy above is Conconi et al. (2012), which examines how domestic fast-track procedures for congressional approval of trade agreements affect the outcome of trade negotiations.

X has some intuitive appeal, but does not stand up to a more rigorous test, which is the following question: can rule X be part of an optimally designed contract, at least for some plausible contracting environment?

I will organize the discussion below as follows. I will first lay out a simple model of TAs where two distinct forms of contract incompleteness—rigidity and discretion—arise endogenously from contracting frictions, and argue that this model can help explain a number of design features of the observed TAs. I will then discuss other models that have been proposed in the literature to explain specific rules such as tariff caps, reciprocity, MFN, market-access rules, and “liability” vs “property” rules. Finally, I will focus on the design of dispute settlement procedures.

3.1.1. Contracting costs, rigidity, and discretion

To lay out some basic concepts I will start by outlining a model by [Horn et al. \(2010\)](#), where the incompleteness of a TA emerges endogenously from the costs of writing a contract. In spite of its simplicity, this form of contracting friction can go surprisingly far in explaining the way TAs are designed.⁴²

At the basis of this model is the observation that there are two important sources of complexity in writing a TA: (a) uncertainty about future economic/political conditions, which calls for agreements that are highly contingent, and (b) the wide array of policies (domestic and border measures) that can affect trade, which in turn calls for agreements that are very comprehensive in their policy coverage. For these reasons, writing a complete contract would be very costly, since all contingencies and policies would need to be described ex-ante and verified ex-post.⁴³ In this context, one can think of two ways to save on writing costs: leaving contingencies out of the contract, which leads to *rigidity*, or leaving policies out of the contract, which amounts to introducing *discretion*.

Real-world TAs exhibit an interesting combination of rigidity and discretion. For example, the GATT-WTO binds trade instruments, but domestic instruments are largely left to discretion, except that (i) the WTO has introduced regulation of domestic subsidies, and (ii) all domestic policies must satisfy the National Treatment rule. Also, constraints on tariffs take the form of ceilings (so governments have downward discretion), and such ceilings are largely rigid, but the contract also provides for “escape clauses” under some contingencies. Horn et al. argue that the presence of contracting costs can help explain these design features.

⁴² Writing costs can be interpreted more broadly as capturing also the costs of negotiating a TA and the costs of verifying contingencies and policies. I also note here that this approach to modeling endogenous contract incompleteness was first developed, though in a different setting, by [Battigalli and Maggi \(2002\)](#).

⁴³ Are contracting costs empirically important for an agreement such as the WTO? This is a legitimate question, but given the huge number of products, countries, policy instruments, and contingencies that are involved in the WTO, and the fact that this agreement took eight years of negotiations to complete, contracting costs seem quantitatively important. Many trade-law scholars agree with this view. For example, [Schwartz and Sykes \(2002\)](#) write: “...Many contracts are negotiated under conditions of considerable complexity and uncertainty, and it is not economical for the parties to specify in advance how they ought to behave under every conceivable contingency.” (pp. 181–182)

Horn et al. consider a two-country, partial-equilibrium model where markets are perfectly competitive but there may be production and consumption (localized) externalities, so that there is an efficiency rationale for multiple policies, in particular tariffs and production subsidies. There can be uncertainty both in the underlying trade volume and in the externality levels, so that a first-best agreement would need to specify policies in a state-contingent way. At the core of the model is the assumption that the cost of the agreement is increasing in the number of policy instruments and contingencies it specifies.

The analysis focuses on four questions of TA design. The first one is whether it is desirable to leave domestic subsidies to the governments' discretion. Horn et al. find that, for a given level of contracting costs, this is more likely if: (a) trade volumes are low, so that countries gain little from manipulating TOT; or (b) countries have little monopoly power in trade, so that they have little capacity to manipulate TOT; or (c) subsidies are not a good substitute for tariffs as a means of manipulating TOT. The trade-volume effect at point (a) suggests an explanation for why domestic subsidies have been constrained by WTO while they were largely left to discretion under GATT, namely, that a general expansion of trade volumes over time has made it more costly to leave subsidies to discretion. Moreover, the "monopoly power" and "instrument substitutability" effects at points (b) and (c) together suggest a possible explanation for why developing countries have been largely exempted from constraints on subsidies through "special and differential treatment" clauses: developing countries typically do not have strong market power and do not have a broad range of domestic policy instruments that can easily substitute for tariffs.

The second design question is whether and how the TA should be state-contingent. An interesting result in this regard is that, conditional on leaving domestic subsidies to discretion, it can be optimal to specify an escape clause that allows a government to increase its tariff when the level of imports is high, as a way to mitigate the stronger incentives to distort domestic subsidies in periods of high import volume. Thus the model suggests a novel explanation for the desirability of escape clauses in TAs: these can be attractive because they provide an indirect means of managing the distortions associated with leaving domestic policies to discretion.⁴⁴

The third point made by Horn et al. is that the presence of contracting costs can explain why tariffs are constrained by *ceilings* rather than by exact levels. More specifically, the optimal agreement may include *rigid* tariff ceilings. Intuitively, conditional on some contingencies being missing from the agreement, leaving downward discretion in tariffs can only be beneficial, since a government is always tempted to distort tariffs *upwards*, and there may be states of the world where the unilaterally optimal tariff lies below the

⁴⁴ Note that this explanation for escape clauses is very different from others that have been proposed in the theoretical literature, and in particular those that are based on self-enforcement considerations (e.g. Bagwell and Staiger, 1990, discussed in Section 3.2.1).

ceiling. In Section 3.1.2, I will discuss other possible explanations for the use of tariff ceilings.

Finally, the model by Horn et al. can provide a novel rationale for the National Treatment (NT) rule, showing that such a rule can serve to save on contracting costs. To make this point, Horn et al. extend the basic model outlined above by allowing for consumption taxes on domestically-produced and imported goods, and formalize the NT rule as a constraint that these taxes be equalized. A preliminary observation is that a TA that imposes the NT rule but leaves discretion over the (non-discriminatory) consumption tax can achieve a new form of discretion that cannot be achieved without the NT rule, namely, discretion over the consumer price wedge; indeed, a non-NT agreement can only leave discretion over the producer price wedge.⁴⁵ Horn et al. then show that under some conditions the NT rule can indeed be part of an optimal TA. The key observation here is that leaving discretion over the (non-discriminatory) consumption tax may involve a subtle benefit, namely that this tax will be responsive to contingencies, thus this form of discretion is an indirect way to make the TA state-contingent. If specifying contingencies is quite costly and the “indirect state-contingency” effect just highlighted is important, then introducing the NT rule in the TA may be optimal.⁴⁶

Before proceeding, I emphasize that this model focuses on a setting where TAs are motivated by TOT externalities, but an interesting and still unexplored question is the extent to which an incomplete-contracting approach of this kind but applied in the context of the domestic-commitment theory might generate new insights and help interpret features of real-world TAs.

3.1.2. Tariff Caps

As I pointed out in the previous section, tariff restrictions in the GATT-WTO take the form of tariff caps. The rationale for the use of tariff caps has been the subject of several

⁴⁵ To understand this slightly cryptic statement, focus on the importer country. In the absence of NT, the wedge between consumer price and world price is given by $p - p^w = \tau + t_f$, where τ is the tariff and t_f is the consumption tax on the imported good, while the wedge between producer price and world price is $q - p^w = \tau + t_f + s - t_h$, where s is the production subsidy and t_h the consumption tax on the domestically-produced good. Next focus on the price wedges under the NT rule, which imposes the constraint $t_h = t_f$. Letting t denote the common level of the internal tax, the NT rule transforms the set of policy instruments from (τ, s, t_h, t_f) to (τ, s, t) , and the price wedges become $p - p^w = \tau + t$ and $q - p^w = \tau + s$. Now notice that with an NT-based agreement that constrains τ and s and leaves t to discretion, it is possible to tie down the producer price wedge $q - p^w$ while leaving discretion over the consumer price wedge $p - p^w$, whereas this is *not* possible with a non-NT agreement.

⁴⁶ The role of the NT rule for domestic taxes has been examined also by Horn (2006). He considers a model where TAs are exogenously incomplete contracts that can include tariff bindings and an NT rule for consumption taxes. Domestic and imported goods are differentiated, so efficiency may call for discriminatory consumption taxes, but Horn shows that, if tariff bindings are set at appropriate levels, including an NT rule always improves global welfare. Saggi and Sara (2008) extend Horn's (2006) analysis by allowing for heterogeneity in product quality and country size. There is also a small literature that examines the role of the NT rule for product standards: see in particular Battigalli and Maggi (2003), Costinot (2008), and Staiger and Sykes (2011).

papers in the literature (in addition to [Horn et al., 2010](#) and [Maggi and Rodriguez-Clare, 2007](#), already discussed above).

[Bagwell and Staiger \(2005a\)](#) consider a model where governments have private information about domestic political-economy shocks, and show that in such a setting tariff caps tend to be preferable to exact tariff commitments. The intuition is similar as in [Horn et al. \(2010\)](#): since the TA cannot be contingent on the political-economy shock, and since a government's temptation is to distort the tariff upward, leaving downward flexibility cannot hurt, and it is strictly preferred if the support of the shock is sufficiently wide. However, just as in [Horn et al. \(2010\)](#), this model can explain only why we do not observe exact tariff commitments, and it stops short of characterizing the *optimal* tariff rule.

[Amador and Bagwell \(2013\)](#) is the first paper that provides a full theoretical explanation of tariff caps, by showing that under some conditions a tariff cap is not only preferable to an exact tariff commitment, but is also the optimal tariff rule. Specifically, Amador and Bagwell consider a partial-equilibrium model where an import tariff is the only available policy instrument, and the importing government has private or non-verifiable information about domestic political pressures. Contingent transfers are not available, thus a TA can only specify a set of permissible tariffs that the importing government may apply. The governments' objective functions are specified in reduced form, say $\Omega(\tau, \gamma)$ for the Home government (where τ is the tariff and γ a political-economy parameter) and $\Omega^*(\tau)$ for the Foreign government.⁴⁷ A tariff cap is shown to be optimal if the convexity of the Foreign objective function ($\Omega_{\tau\tau}^*$) is not too pronounced relative to the concavity of the Home objective function ($\Omega_{\tau\tau}$) and the density of γ does not decrease too fast. Amador and Bagwell then examine when these conditions are satisfied in the context of two specific market structures, a perfectly competitive setting and a monopolistic competition setting.

In reality, some governments do exercise the downward discretion afforded by tariff caps and apply tariffs strictly below the cap levels (the so-called “binding overhang”). Empirically there is considerable variation in the extent of binding overhang, as well as in the levels of tariff ceilings, across countries and sectors (see for example [Bacchetta and Piermartini, 2011](#)). [Beshkar et al. \(2011\)](#) propose a model that sheds light on one important dimension of this variation. They consider a model where governments have private information about domestic political pressures, and examine how the levels of tariff ceilings and the expected size of the binding overhang depend on countries' market power. Their main result is that, when a country has more market power, the optimal tariff ceiling is lower and the expected binding overhang is smaller. This is a consequence of the fact that when a country has more market power its tariffs exert stronger TOT externalities, so providing flexibility through higher tariff bindings causes greater efficiency

⁴⁷ I note here that [Amador and Bagwell \(2013\)](#) consider a slightly more general setting than the one described in the text, where the importing government is allowed to “burn” money.

loss. Beshkar et al. then present econometric evidence in support of this prediction, using a dataset on applied and bound tariffs for WTO member countries.⁴⁸

It is useful at this juncture to recall from Section 2.2 that there is another possible explanation for the use of tariff caps, which is not based on frictions in government-to-government contracting, but rather on domestic-commitment issues, as pointed out by Maggi and Rodriguez-Clare (2007). One interesting difference between these two explanations of tariff ceilings—the one based on international contracting frictions and the one based on domestic-commitment issues—is that the former predicts binding overhang with positive probability in equilibrium, whereas the latter predicts no binding overhang in equilibrium. For this reason, while domestic-commitment considerations may be part of the explanation for tariff ceilings, they are probably not the whole story behind them.⁴⁹

One of the challenges of this body of theory is to explain not just why the WTO imposes rule X or why it imposes rule Y, but also why it imposes rules X and Y. This is a non-trivial question, since different rules can interact in complex ways, and often it is not easy to rationalize the WTO's joint use of disparate approaches to regulating trade policies. Beshkar and Bond (2012) take a step in this direction by developing a model that can explain why the GATT-WTO combines two different forms of flexibility on tariffs, namely tariff caps and escape clauses. More specifically, the model has a similar structure as Amador and Bagwell (2013), except that a government can produce evidence (at a cost) about the state of its domestic political pressures. Beshkar and Bond show that it may be desirable to combine a tariff cap with an escape clause that allows a government to raise the tariff above the cap if it produces evidence that doing so is politically efficient. They also show that these two forms of flexibility are substitutes, in the sense that the optimal tariff ceiling is lower—and tariff overhang may even disappear—in the presence of the escape clause than in its absence.

3.1.3. “Shallow” vs “Deep” Integration

The GATT agreement was largely based on a “shallow integration” approach, in the sense that the agreement placed direct constraints on border measures (such as trade taxes and quotas), while domestic policies were largely left to discretion, except for the requirement that they not be used to erode the market-access levels previously nego-

⁴⁸ Amador and Bagwell (2012) also presents an interesting result regarding binding overhang. They consider a variant of Amador and Bagwell (2013) where governments have private information about the value of tariff revenue, a setting that is arguably relevant for developing countries. In a linear-quadratic specification with a uniform type distribution, they show that the optimal tariff cap and the probability of binding overhang are higher when there is greater uncertainty in the type distribution and when the upper bound of the support of the distribution is higher.

⁴⁹ It is also relevant to note however that, if tariff ceilings were due solely to the presence of non-contractible contingencies, we would expect that, for a given product in a given country, the applied tariff is sometimes at the bound level and sometimes below it, but it is not clear from existing evidence that this is actually the case in reality.

tiated by governments.⁵⁰ Bagwell and Staiger have argued in several papers that the GATT’s “non-violation” clause (Art. XXIII.1b) can be interpreted as imposing a kind of “market-access-preservation” constraint on governments. The WTO, on the other hand, has gone beyond a shallow integration approach, by introducing direct restrictions on some domestic policies, notably production subsidies, so in this sense it has moved closer to a “deep integration” approach. What are the relative merits of these two approaches, and why has the approach changed in going from GATT to WTO?

The benchmark paper on this topic is [Bagwell and Staiger \(2001\)](#), which focuses on a standard two-good model with perfect competition and no uncertainty. I will briefly discuss this paper using the notation introduced in Section 2.1.2. The basic idea can be illustrated by supposing that the Foreign government is passive, so its objective can be written simply as $\Omega^*(p^w)$, whereas the Home government can use a tariff τ and a non-wedge policy v (e.g. a labor standard), so that its objective can be written as $\Omega(v, p, p^w)$. Supposing that Home imports the non-numeraire good, let $X^*(p^w)$ be Foreign’s export supply for this good and let $p^w(\tau, v)$ be the equilibrium world-price level. In its bare-bone form, a “shallow integration” agreement specifies a reference tariff level τ^A and allows Home to choose its policies (including the tariff) subject to the only constraint that the resulting trade volume be the same as that implied by τ^A and the noncooperative domestic policy level, say v^N . Such a market-access-preservation constraint thus takes the form: $X^*(p^w(\tau, v)) = X^*(p^w(\tau^A, v^N))$.⁵¹

It is immediate to show that, under perfect competition and in the absence of uncertainty, a shallow-integration agreement is sufficient to achieve global efficiency. The intuition is straightforward: since there is a one-to-one link between trade volume and TOT, a trade-volume-preservation constraint effectively prevents the Home government from manipulating TOT; and since TOT manipulation is the only cause of inefficiency in the Nash equilibrium, this constraint is sufficient to achieve efficiency.⁵² This argument can be extended to allow the Foreign government to be policy-active as well.⁵³

⁵⁰ In addition, GATT imposed that certain domestic instruments such as consumption taxes and product standards must satisfy the National Treatment rule, as I discussed in Section 3.1.1.

⁵¹ Note that a simpler but equivalent contract would be one that specifies a trade volume \bar{X}^* that Home is required to deliver, so Home chooses (τ, v) subject to the constraint $X^*(p^w(\tau, v)) = \bar{X}^*$. Clearly this contract is equivalent to the one mentioned in the text, because choosing the baseline tariff level τ^A effectively amounts to choosing the target trade level \bar{X}^* .

⁵² See also [Bagwell and Staiger \(2006\)](#), who extend the analysis to a setting where governments can use trade policies and production subsidies. Another related paper is the one by [Bajona and Ederington \(2012\)](#), who argue that, if the TA must be self-enforcing and domestic policies are not observable, the optimal TA may include not only market-access constraints but also tariff bindings.

⁵³ There is an important caveat to this argument: in GATT-WTO the negotiated tariff levels are not simply reference levels, but tariff *ceilings*. In [Bagwell and Staiger’s \(2001\)](#) model, if the contract includes also a tariff ceiling together with the market-access constraint, the inability of a government to raise the tariff above the ceiling may under some conditions lead to inefficient outcomes. This observation leads Bagwell and Staiger to argue that WTO rules could be made more efficient while at the same time affording governments more sovereignty.

Bagwell and Staiger's (2001) result represents an important benchmark, but a number of subsequent papers have highlighted reasons why a shallow integration approach may *not* be sufficient to achieve globally efficient outcomes. One reason may be the presence of non-TOT externalities, such as the ones emphasized by the New Trade theories; a second reason may be the presence of domestic-commitment motives for TAs; and a third reason may be the presence of contracting frictions.

A setting with non-TOT externalities where shallow integration is not sufficient to achieve efficiency is the one considered by Antras and Staiger (2012a,b). What is responsible for the insufficiency of shallow integration in that setting is that prices are determined by bilateral bargaining rather than market clearing. This is intuitive in light of the fact that, as discussed in Section 2.3 above, the purpose of a TA in these models goes beyond the correction of TOT externalities.⁵⁴ Another example is provided by DeRemer (2011), who argues that shallow integration is not sufficient to achieve efficiency in a monopolistic competition model with delocation externalities, when governments can use trade taxes and production subsidies. Moreover, a natural conjecture is that the same would be true in a Cournot setting with profit-shifting externalities, such as the one in Mrazova (2011) or Bagwell and Staiger (2012a), if domestic policies were allowed in these models.

It should be clear from the discussion above that a shallow integration approach can achieve efficiency only if the international policy externalities travel solely through the TOT channel. It should not come as a surprise, therefore, that if the TA is motivated by domestic-commitment motives then a deep integration approach will be required to achieve efficiency. This point is made in the paper by Brou and Ruta (2009), which I already mentioned in Section 2.2. One important point that emerges clearly from all the papers discussed above is that *the motives for a TA crucially affect the optimal design of a TA*.

As I mentioned earlier, a final reason why a shallow integration approach may not be enough to induce globally efficient outcomes is the presence of contracting frictions. To understand why, let us go back to the Bagwell–Staiger setting considered at the beginning of this section, with perfect competition and a single policy-active country (Home). Suppose that the state of the world is uncertain at the time the TA is negotiated, but Home observes it before choosing its policies. Then, in order for a market-access-preservation rule to induce a globally efficient outcome, such rule would need to be state-contingent. But if some contingencies are not verifiable (or if specifying them in the contract is costly), then a shallow integration approach may not achieve full efficiency. This point is made by Lee (2007, 2011) in the context of a model where governments choose trade taxes and production taxes, and have private information about the level of a domestic externality.

⁵⁴ However it is important to note that the two questions, “Are the PO policies efficient?” and “Is shallow integration sufficient to achieve efficiency?” do not necessarily yield the same answer. It is possible that PO policies are efficient but shallow integration does not achieve efficiency: in Antras and Staiger (2012a), for example, in the case of welfare-maximizing governments PO policies are efficient but shallow integration does not achieve efficiency.

I now feel the need to take a step back and reflect on the way this design question has been addressed in the literature, and on the way I think it should ideally be framed. The typical approach in the literature (at least implicitly) has been to ask whether a shallow-integration agreement can achieve global efficiency: if the answer is yes, then a shallow-integration agreement is considered preferable to a deep-integration agreement, based on an (implicit or explicit) assumption that “more sovereignty is always better;” and if the answer is no, the conclusion is that a deep-integration agreement is preferable. I find this conceptual approach useful but not entirely satisfactory, for the reasons I explain next.

The argument that it is preferable, other things equal, to give countries sovereignty over domestic policies is presumably based on a notion that relinquishing control over domestic policies entails some cost to a government. This idea has some intuitive appeal, but it is not obvious how it can be given a rational-choice justification. Why would governments not feel similarly about relinquishing control over border measures? Perhaps one way to rationalize the “sovereignty” argument is that domestic policies are needed to address domestic economic or political needs that vary over time, and taking away sovereignty over domestic policies removes a government’s flexibility in responding to these varying needs. But this implicitly assumes that the TA constrains domestic policies in a *rigid* way; in the absence of contracting frictions, a TA would constrain these policies in a contingent way, so there would be no shortcoming from a deep-integration approach. Thus one *can* rationalize the idea that giving up sovereignty is costly, but *at the core*, this boils down to an argument that contracting frictions cause rigidity in TAs. For this reason, I think that introducing uncertainty and contracting frictions explicitly in our conceptual framework is important to examine the tradeoff between shallow and deep integration.

The discussion above in turn suggests an alternative, and I think preferable, approach to the comparison between shallow and deep integration. The key question in my opinion should be: what are the transaction costs associated with each of these two approaches, and which one is more efficient when transaction costs are taken into account? The answer to this question is far from obvious: a deep-integration agreement requires specifying all border and domestic policies, and this is likely to involve large transaction costs; a shallow-integration agreement, on the other hand, may save on the costs of contracting over domestic policies, but it will achieve less efficient outcomes (if the environment is uncertain); and if specifying contingencies in the TA is costly, the tradeoff becomes even more complicated. If one takes this explicit transaction costs perspective, then the results of the existing literature appear as a valuable step, but only a first step.

3.1.4. Reciprocity and MFN

The principles of reciprocity and MFN are two major pillars of the GATT-WTO. The role of these principles has been examined by Bagwell and Staiger in a number of writings,

and most notably in the context of the perfect-competition setting of [Bagwell and Staiger \(1999a\)](#). I will start by focusing on the principle of reciprocity in the context of a two-country model (where MFN of course plays no role).

Bagwell and Staiger define a reciprocal change in trade policies as one that leaves world prices unchanged.⁵⁵ More specifically, they distinguish between two notions of reciprocity. The first one is a principle guiding tariff *negotiations* starting from the Nash equilibrium: Bagwell and Staiger show that, if tariff negotiations satisfy reciprocity (that is, if they leave world prices unchanged relative to the Nash equilibrium), they will lead to a Pareto-improvement over the Nash equilibrium, although in general they will not lead all the way to the Pareto-efficient frontier.⁵⁶ An important note of interpretation is the following: in the GATT-WTO the principle of reciprocity as it applies to tariff negotiations is not a strict rule, but just an informal principle, so [Bagwell and Staiger's \(1999a\)](#) analysis of reciprocity-guided negotiations is best interpreted in normative terms, as highlighting how the negotiation outcome would be affected *if* reciprocity were strictly imposed as a rule.

The second notion of reciprocity considered by Bagwell and Staiger applies to tariff *renegotiations*. Unlike the informal reciprocity principle that applies to negotiations, GATT-WTO does impose a formal rule of reciprocity for tariff renegotiations, specifically in Article XXVIII of GATT. Bagwell and Staiger formalize this rule through a two-stage negotiation game, as follows: (i) in the first stage, governments negotiate a pair of tariffs, with the disagreement point given by the Nash equilibrium tariffs; (ii) in the second stage, governments can renegotiate the tariffs, and if the renegotiation fails, a government can unilaterally change its tariff from the level that was agreed upon in the first stage, but at the condition that the trading partner get compensated through a reciprocal tariff change.

Before proceeding, I make an observation that will be useful later on. Notice the nature of the *disagreement point* for the process of renegotiation under reciprocity: if the renegotiation fails, a government can choose to unilaterally “breach” the contract and compensate the trading partner, with the compensation taking the form of a reciprocal tariff increase by the trading partner. This observation leads me to note that there is a simpler way to interpret such reciprocity rule, which I find more illuminating. Rather than imposing a constraint on renegotiation, this rule can equivalently be modeled as changing the nature of the contract itself, from one that specifies tariff commitments without allowing for “breach” (in law and economics jargon, a “property” contract), to one that allows a government to breach the contract in exchange for a certain amount of “damages” (in law and economics jargon, a “liability” contract). Notice that, if reciprocity is modeled in this way, simply as a liability rule incorporated in the initial contract, then there is no scope for renegotiation in Bagwell and Staiger’s setting, and we can simply think

⁵⁵ As Bagwell and Staiger show, an equivalent definition is that changes in trade policies are reciprocal if they bring about equal-value changes in each country’s volumes of imports and exports when valued at the initial world prices.

⁵⁶ The only case in which negotiations according to reciprocity lead to Pareto efficiency is the one where countries are perfectly symmetric.

of governments as negotiating a contract within this particular class (liability contracts with breach damages given by reciprocal tariff changes), ignoring renegotiation.⁵⁷

Bagwell and Staiger's main results regarding the rule of reciprocity are two. The first one is that the only efficient tariff pair that can be implemented under reciprocity (or, adopting my interpretation above, when the contract is restricted to be a liability contract with breach damages given by reciprocal tariff changes) is given by the politically efficient (PO) tariffs. This result is a consequence of the fact that, when viewed in tariff space, each government's iso-payoff curve is tangent to the iso-world-price curve at the PO point, thus starting from this point there is no unilateral incentive for a government to move along the iso-world-price curve. This in turn implies that the reciprocity-constrained Pareto frontier lies inside the unconstrained Pareto frontier, except at the PO point. The second result is that, when governments bargain over the reciprocity-constrained Pareto frontier, the outcome is generically inefficient (it is not the PO point), but it tends to be closer to the PO point as compared with the unrestricted bargaining scenario.⁵⁸

Bagwell and Staiger interpret these results as suggesting that the reciprocity rule induces a rebalancing of power across countries, since it moves the negotiation outcome toward a point that is not affected by bargaining powers (the PO point). This conclusion resonates with the GATT's emphasis on "rules" vs "power," but as remarked above, the rebalancing of power induced by the reciprocity rule in general entails an "efficiency penalty." Bagwell and Staiger then argue informally that reciprocity may provide efficiency gains if it encourages weaker countries to participate in GATT. This idea is based on McLaren's (1997) model: recall from Section 2.2 that in this model a small country trading with a large country may prefer to stay out of a TA to avoid being "held up," and for this reason the large country would like to commit *not* to exploit its bargaining power, in order to encourage the small country to participate in the TA. Thus the broad idea is that, if one thinks of GATT as initially including a set of large/powerful countries, but there is also a set of smaller/weaker countries that may consider accessing GATT at a later stage, the initial members may prefer to commit not to exploit their bargaining power in future negotiations, in order to encourage other countries to seek participation.

I will make two further comments about Bagwell and Staiger's analysis of the reciprocity rule. The first one is a "devil's advocate" comment. Bagwell and Staiger's interpretation of the results outlined above is that reciprocity "works well" when TOT externalities are the only motive for a TA. But one might argue that the model tells the opposite story: reciprocity causes the negotiation outcome to be inside the Pareto frontier, so the world would be more efficient *without* reciprocity. A more cautious interpretation of Bagwell and Staiger's analysis would be that it provides a positive (as opposed to normative)

⁵⁷ I note that this way of thinking about reciprocity, as a rule specifying breach remedies rather than as a constraint on renegotiation, is close to the way Ossa (2011) formalizes the reciprocity rule in his model.

⁵⁸ I say "tends to be closer to the PO point" because this is true only under some conditions, as explained in footnote 25 of Bagwell and Staiger (1999a).

evaluation of the reciprocity rule, highlighting that this rule has a distributional effect (which might be desirable in a richer model that includes participation considerations) and an efficiency cost.

The second comment is that, aside from the participation argument outlined above, Bagwell and Staiger's analysis is suggestive of another potential efficiency benefit of reciprocity. Since, as I argued above, imposing the reciprocity rule is akin to structuring the TA as a (specific type of) "liability" contract, this might be appealing as a way to inject flexibility in the TA when countries are subject to shocks and the TA cannot be fully contingent. But this is just suggestive, since there is no uncertainty or contract incompleteness in the model.⁵⁹ This brings me back to one of my overarching points, namely that a better understanding of TA design requires bringing transaction costs explicitly into the picture.

After their analysis of reciprocity within a two-country setting, Bagwell and Staiger (1999a) turn to an examination of the implications of the MFN rule within a multi-country setting. As I mentioned in Section 2.1.2, in a multilateral world where governments can set discriminatory tariffs, there is a whole vector of bilateral offshore prices, which generates a complicated pattern of international policy externalities, but the MFN rule has the effect of channeling all these externalities into a single world-price externality. Building on this observation, Bagwell and Staiger show that the MFN rule, if used in tandem with the reciprocity rule, guides countries toward the PO tariffs (which, recall, are efficient under MFN). The key point here is that the combination of MFN and reciprocity has similar effects in a multi-country world as the reciprocity rule does in a two-country world.

So far I have focused on the implications of reciprocity and MFN within a perfectly competitive environment. It has been argued, however, that reciprocity and MFN can be rationalized also within an imperfectly competitive environment, where the motives for a TA go beyond the correction of TOT externalities. In particular, Ossa (2011) argues that, in a monopolistic-competition setting, reciprocity and MFN can be interpreted as helping countries internalize the production-delocation externalities generated by trade policies.⁶⁰

3.1.5. "Property" vs "Liability" Rules

In the previous section I mentioned the distinction between a "property" contract (one that does not provide for the possibility of breach) and a "liability" contract (one that

⁵⁹ As I will discuss in the next section, papers by Maggi and Staiger (2012) and Beshkar (2010a,b) have examined this idea more formally in models with non-verifiable or privately observed shocks.

⁶⁰ It should be noted however that the implications of reciprocity and MFN in Ossa's model are somewhat different than in Bagwell and Staiger's model. Ossa shows that reciprocity can help countries achieve an efficient outcome if reciprocity is applied to multilateral trade negotiations, and then argues that MFN can serve to multilateralize trade negotiations; Bagwell and Staiger, on the other hand, show that reciprocity and MFN ensure that any bilateral negotiation will lead to an efficient outcome.

gives a government the option to breach-and-compensate). In real-world TAs, there is considerable variation between liability-type rules and property-type rules, both across issues and over time. For example, [Pauwelyn \(2008\)](#) argues that property rules are the “default” approach in the WTO and NAFTA, but for certain issues such as tariff bindings and production subsidies, a liability-rule approach has been taken.⁶¹ Moreover, most legal scholars take the view that the early GATT operated as a system of liability rules, while in more recent times the GATT/WTO has evolved toward a property-rule system (see for example [Jackson, 1997](#)).⁶²

As mentioned above, a liability approach is appealing in the presence of uncertainty because it can inject flexibility in the TA without the need to specify contingencies explicitly in the contract: intuitively, a liability rule can induce a country to internalize the externalities that its trade policy exerts on its trading partners. However, in the international trade arena, the liability approach has an important shortcoming, namely that government-to-government compensation is inefficient: cash transfers are typically not available, and compensation typically takes the form of “self-help” through tariff retaliation. For this reason, a liability rule can generate deadweight loss, and this gives rise to a non-trivial tradeoff between property and liability rules.⁶³

The tradeoff between property and liability rules becomes even more subtle if governments can renegotiate the TA, because in this case a property rule does not necessarily imply inflexible policy outcomes (since it can be renegotiated ex-post), and it can give rise to inefficient compensation in case of renegotiation. Taking renegotiation into account seems important also in light of its empirical importance in the context of GATT-WTO, where renegotiations have taken place in many instances over the years.

The choice between property and liability rules in the presence of renegotiation is analyzed by [Maggi and Staiger \(2012\)](#). In this model, governments negotiate over a binary trade policy (free trade or protection) under uncertainty about the future joint benefits of protection (which can be positive or negative, due to political economy considerations), and can renegotiate the TA after the uncertainty is resolved. A key feature of the model, in line with the discussion above, is that government-to-government compensation entails a deadweight cost.

⁶¹ This includes for example the provisions for escape from negotiated tariff bindings (GATT Articles XIX and XXVIII, respectively), the rules on “actionable” production subsidies in WTO, and the provisions to protect investors against expropriation in NAFTA (and in many other bilateral investment treaties).

⁶² Here I note that, while the terminology of property and liability rules is more common in the law-and-economics literature, the choice between these two contractual forms is an important topic in the economic literature on optimal contract design, where a liability contract is often referred to as an “option contract,” and a property contract is often referred to as a “noncontingent contract;” or simply a “property-right” contract (see for example [Segal and Whinston, 2002](#)).

⁶³ The inefficiency of government-to-government compensation is not the only possible shortcoming of a liability approach. Another limitation—which I abstract from here—is that the damage inflicted by trade protection on trading partners is typically non-verifiable, and this makes it impossible to induce a government to perfectly internalize trade policy externalities.

Maggi and Staiger find that a property rule is optimal if uncertainty about the joint benefits of protection is small enough, while a liability rule is optimal when this uncertainty is large. If one interprets uncertainty about the joint benefits of protection as due to political–economy shocks, then this result suggests that liability rules should be more prevalent in issue areas that are more politically sensitive. This prediction seems consistent with the observation that the WTO has taken a liability approach in the areas of import tariffs and production subsidies (which are arguably very sensitive to political–economy shocks) and a property–rule approach in other areas. A further result is that, if a liability rule is optimal, the optimal level of damages falls short of fully compensating the exporter, contrary to the well-known “efficient breach” argument. This result is shown also in a related model by [Beshkar \(2010a\)](#), though in a setting without renegotiation.⁶⁴

The model yields predictions also about the occurrence of renegotiation in equilibrium, and how it correlates with the optimal contract form. I report two findings here. The first is that renegotiation, when it occurs, always entails the exporter agreeing to compensate the importer in exchange for trade liberalization.⁶⁵ The second is that an optimal property rule is never renegotiated in equilibrium. This finding may seem counterintuitive, since a property rule is inherently rigid and renegotiation should be useful to mitigate such rigidity, but it turns out that renegotiation can improve the performance of a property rule only if such rule is suboptimally adopted.⁶⁶

3.1.6. *Dispute Settlement Procedures and Contract Incompleteness*

As I discussed at the beginning of Section 3, the role of the WTO’s DSB seems to go well beyond a pure enforcement role, at least judging from casual observations of real-world trade disputes, where the DSB seems to often play an “activist” role by interpreting vague clauses and filling gaps of the agreement. The potential role of the DSB in completing an incomplete agreement is the focus of a paper by [Maggi and Staiger \(2011\)](#).⁶⁷

⁶⁴ [Beshkar \(2010b\)](#) builds on [Beshkar \(2010a\)](#) by allowing the WTO court (DSB) to observe a noisy signal of the state of the world, and shows that the performance of the contract can be improved by making the remedies for breach (that is, the injured country’s retaliatory tariff) contingent on the DSB’s signal, which in turn can be accomplished by allowing the injured country to file a complaint that triggers DSB intervention. See also [Beshkar \(2011\)](#), which extends the previous models by allowing for a limited form of renegotiation between governments.

⁶⁵ This asymmetry in the predicted direction of renegotiation is a consequence of two features: that the contractual obligation is to liberalize trade, and that it is never optimal to induce renegotiation in states of the world where the threat point is the contractual obligation itself.

⁶⁶ In the model just discussed, the court/DSB does not play an active role in equilibrium, so the model is silent about what determines the occurrence of trade *disputes* and their outcomes. In a more recent working paper, [Maggi and Staiger \(2013\)](#) consider a richer model in which trade disputes can occur in equilibrium and can result in a variety of outcomes; in particular, governments may settle early or trigger a DSB ruling, and in the latter case, they may implement the DSB ruling or renegotiate after the ruling. Two new features of the model are responsible for this rich set of possibilities: first, the DSB can observe a noisy signal of the joint benefits of protection (interpreted as the outcome of a DSB investigation), so governments are uncertain about the direction of the DSB ruling; and second, governments have a further opportunity to renegotiate the TA *after* the DSB issues a ruling.

⁶⁷ An early attempt to examine the potential role of the DSB in completing an incomplete contract can be found in [Battigalli and Maggi \(2003\)](#), who focus on agreements on product standards. In that model, the TA cannot specify

Maggi and Staiger consider three possible activist roles for the DSB: *interpreting* vaguely stated obligations; *filling gaps* in the agreement; and *modifying* rigid obligations; and for each of these roles, the DSB may or may not have authority to set *precedent* for future rulings. Governments design the institution—that is the combination of contract and DSB mandate—under uncertainty about the future state of the world. The relevant contingencies are assumed to be too costly to describe in a crisp way, so the contract is necessarily incomplete. The model allows for three possible forms of contractual incompleteness: the first two, *rigidity* and *discretion*, are familiar (see Section 3.1.1); the third one, *vagueness*, is novel.⁶⁸ There is a natural pairing between these three forms of contractual incompleteness and the three possible activist DSB roles described above: the DSB can interpret a vague contract; it can fill gaps if the contract leaves discretion; and it can grant exceptions if the contract is rigid. Or, the DSB can play a non-activist role and simply enforce clearly stated obligations. Furthermore, for each of the activist roles, the DSB may or may not have precedent-setting authority. A key feature of the model is that invoking the DSB entails two kinds of inefficiency: one is that the governments incur litigation costs, and the other is that DSB rulings are imperfectly accurate.

Maggi and Staiger show that the optimal institutional form depends critically on the degree of DSB accuracy. If the DSB is sufficiently accurate, it is optimal to leave gaps in the contract and give the DSB a mandate to fill those gaps. If the DSB is sufficiently *inaccurate*, it is optimal to write a vague or rigid contract and give a non-activist mandate to the DSB. And if the level of DSB accuracy is intermediate, it is optimal to write a vague contract and allow the DSB to interpret the contract. The “modification” role of the DSB turns out to be always suboptimal.

The model delivers good news and bad news regarding the potential of the DSB to enhance efficiency. The good news is that, if the DSB is sufficiently accurate, the first-best outcome can be achieved, in spite of the incompleteness of the contract, the costs of using the DSB, and the imperfection in DSB rulings. The reason is that the threat of invoking the DSB and the expectation of a not-too-inaccurate DSB ruling are enough to discourage opportunistic behavior by governments. This suggests that an activist DSB can generate substantial efficiency gains even if its information is not perfect. The bad news is that the outcome tends to be efficient only when the DSB is *not* invoked in equilibrium, and disputes are more frequent when the DSB is *less* accurate. Thus, the

standards for products that will appear in the future. If a new product appears, governments can negotiate ex-post over the standards for that product, but international transfers are not available ex-post, and as a consequence the ex-post negotiation outcome is ex-ante inefficient. The DSB acts as an “arbitrator” that can be invoked if the ex-post negotiation fails. Battigalli and Maggi show that the DSB can improve ex-ante efficiency by specifying ex-post—if invoked—the standards that should apply to the new product, provided the DSB has sufficiently precise information.

⁶⁸ Vagueness is modeled as a language whose meaning is only partially defined. As an example, consider a contract stating that trade protection is allowed only if “there is substantial injury to the domestic industry.” The idea is that there are states of the world where the latter sentence is clearly true (e.g. if there is an import surge, the domestic industry shuts down, and the majority of workers in the industry are unemployed), others where it is clearly false (if none of the above events has occurred), and there are “gray area” states where it is not defined whether the sentence is true or false.

efficiency-enhancing effect of the DSB is associated with its *off-equilibrium* impacts. If the DSB is invoked in equilibrium, it is always because one of the governments is being opportunistic: either the importer is protecting when it should not, hoping that the DSB ruling will get it wrong; or the exporter is trying to force free trade by filing a dispute when it should not. A corollary of these observations is that the frequency of DSB use is not a good indicator of the performance of the institution.

The model also has interesting implications regarding the “bias” in observed DSB rulings. Because of selection effects in the filing of disputes, DSB rulings tend to have a pro-trade bias if litigation costs fall more on the exporter government than on the importer government. In reality, it is arguably the case that litigation costs fall more on the exporter government, because the burden of proof falls on the complainant, and the complainant is typically the exporter government. Thus the model suggests a possible explanation for the fact that, both under the GATT and the WTO, complainants have mostly won their cases.

Finally, Maggi and Staiger extend the basic model to a two-period setting in order to examine whether it is desirable to give the DSB precedent-setting authority. This is an important issue of institutional design, and one that is receiving increasing attention from legal scholars.⁶⁹ Introducing precedent in this setting is shown to have two opposite effects on efficiency. The beneficial effect of precedent is that it reduces the probability of dispute occurrence *tomorrow*, by clarifying the obligations that will apply should the same state of the world occur again; this leads to a savings in litigation costs. The harmful effect of precedent is that it increases the probability of dispute occurrence *today*; this in turn implies more waste in litigation costs and a less efficient policy outcome (because the DSB ruling is subject to error). Maggi and Staiger show that, as a net result of these effects, precedent is more likely to be beneficial when the accuracy of DSB rulings is lower and when governments care less about the future, or are less likely to interact repeatedly.

3.2. Imperfect Enforcement and Trade Agreements

In this section I will focus on how the presence of enforcement frictions can shape the design of rules and procedures in a TA. More specifically, I will discuss how the presence of self-enforcement constraints affects the design of substantive policy rules, of enforcement rules (i.e. rules that regulate punishments), and of dispute settlement procedures.

In the literature, the dominant approach for modeling a self-enforcing TA has been to consider a game where governments choose trade policies repeatedly and focus on (constrained-)Pareto-efficient equilibria of this game. The implicit assumption in this approach is that governments bargain efficiently over the set of equilibria of the repeated

⁶⁹ Even at the positive level, the extent to which the WTO-DSB currently operates on a precedent system seems subject to debate. According to Jackson (2006, p. 177), “there is quite a powerful precedent effect in the jurisprudence of the WTO, but ... it is not so powerful as to require panels or the Appellate Body considering new cases to follow prior cases.” Jackson concludes that “the ‘flavor’ of the precedent effect in the WTO is still somewhat fluid.”

game. This approach has become fairly standard in the literature and is explained in [Staiger's \(1995a\)](#) chapter, so I will take it for granted and simply provide an informal overview of the contributions in this area after 1995.⁷⁰

3.2.1. Policy Rules

The presence of self-enforcement constraints can have deep implications for the design of policy rules. This point was first made in a pathbreaking paper by [Bagwell and Staiger \(1990\)](#), which showed that, in the presence of (publicly observed) i.i.d. shocks to trade volume, the need to accommodate self-enforcement constraints makes it desirable to include an *escape clause* in the TA. The basic idea is that, in periods of high import volume, a country has a stronger incentive to deviate from its trade policy commitments, so in such periods it may be a good idea to allow a country to escape from its commitments in order to keep cooperation sustainable.⁷¹

More recently, [Bagwell and Staiger \(2003\)](#) have extended their previous model by allowing for persistent shocks to trade volume as well as an acyclic component of the shock. In this setting, [Bagwell and Staiger](#) show that trade protection tends to be countercyclical. The reason is that a boom phase tends to be characterized by fast growth in trade volume, which helps countries sustain lower tariffs than in a recession phase, while acyclic increases in trade volume give rise to increases in tariffs, for a similar reason as in [Bagwell and Staiger \(1990\)](#).⁷²

Another group of papers has highlighted that the presence of self-enforcement constraints can help explain why TAs typically take a *gradual* approach to trade liberalization. The first two papers in this group are [Staiger \(1995b\)](#) and [Devereux \(1997\)](#), the former focusing on the implications of sector-specific skills that depreciate when not in use, and the latter on the implications of technological learning-by-doing. [Furusawa and Lai \(1999\)](#) and [Chisik \(2003\)](#) propose two further mechanisms that can generate gradual trade liberalization as part of an optimal self-enforcing TA: [Furusawa and Lai](#) focus on the role of adjustment costs in worker reallocation across sectors, and [Chisik](#) focuses on the role of irreversible investments in country-specific export capacity. The common theme in all of these papers is the non-stationary nature of the repeated game between governments,

⁷⁰ Before surveying the relevant literature, I mention briefly four papers that do not focus on issues of rule design, but on a more classic question: what conditions facilitate the self-enforcement of TAs? [Furusawa \(1999\)](#) examines how the sustainability of cooperation is affected by the governments' relative patience and the lag between detection of a violation and retaliation; [Park \(2000\)](#) focuses on a trade agreement between a small country and a large country, and examines how the set of sustainable payoffs is affected by the availability of direct transfers and the presence of sunk investments; [Conconi and Sahuguet \(2009\)](#) explore the impact of policymakers' horizons and alternative electoral regimes on the sustainability of international cooperation; and [Conconi and Perroni \(2009\)](#) examine whether and how the ability of governments to commit to policies in the domestic arena affects the sustainability of international cooperation.

⁷¹ See also [Milner and Rosendorff \(2001\)](#) for a related model of self-enforcing agreements where the presence of uncertainty makes it optimal to introduce escape clauses.

⁷² The question of whether or not trade barriers are countercyclical has been the subject of interesting empirical work recently. See in particular [Bown and Crowley \(2013b\)](#) and [Rose \(2013\)](#).

whereby an initial reduction in tariffs leads to a change in some economy-wide state variable (such as the level of physical or human capital allocated to the export sector), which in turn relaxes the self-enforcement constraint and allows governments to sustain further tariff reductions.

One might think that gradual trade liberalization can *only* be explained in a non-stationary trading environment, but [Bond and Park \(2004\)](#) make the surprising point that gradualism can arise even in a stationary economic environment. Specifically, they show that the optimal self-enforcing TA may entail gradual tariff reductions if countries are asymmetric. In this case, it is possible that in the initial phase of the agreement the self-enforcement constraint of only one country is binding, and given this initial asymmetry, the most efficient way to provide incentives to such country is to “backload” its payoff, that is, to promise this country a rising payoff over time.⁷³

3.2.2. Enforcement Rules

When countries negotiate a TA, they negotiate not only substantive policy rules, but also rules that regulate punishments for violations of the agreement. I will refer to these as “enforcement rules.”

A key consideration for the design of enforcement rules is that they must themselves be self-enforcing, or in other words they must be *credible*: in the language of repeated-game theory, this means that the punishment strategy must be an equilibrium of the continuation game after the initial deviation. In this perspective, the question “What are the optimal enforcement rules?” can be formally phrased as “What is the optimal equilibrium punishment strategy?” It is important to keep in mind that in a repeated game there is a vast multiplicity of equilibrium punishment strategies, just as there is a vast multiplicity of overall equilibria. The implicit assumption in this modeling approach is that, when governments negotiate a TA, they bargain efficiently over policy rules *and* enforcement rules, subject to the constraint that all rules be part of an overall equilibrium of the repeated game.

[Maggi \(1999\)](#) examines the optimal design of enforcement rules in the context of a multilateral trading system. A key question in this context is whether and to what extent punishments should be multilateral rather than bilateral. To be concrete, the question is: if country A cheats on country B, should country C be involved in the punishment? Consider first a benchmark scenario where bilateral trading relationships are symmetric and separable (in the sense that changing a bilateral trade barrier does not affect third countries): in such scenario, Maggi shows that there are no gains from multilateral punishments. The intuition is that, while making punishments multilateral increases the future loss from

⁷³ Another paper that highlights implications of self-enforcement constraints for the design of policy rules is [Mrázová \(2011\)](#). This paper (which I already mentioned in Section 2.3.1) argues that the need to make the TA self-enforcing, in conjunction with the presence of profit-shifting externalities and costs of administering policy instruments, can help explain the WTO ban on export subsidies.

a deviation, it also increases the one-time gain from deviating, because if a country is to deviate it will do so against all trading partners; when bilateral relationships are symmetric and separable these two effects cancel each other out. Next consider a scenario characterized by bilateral imbalances of power, in the sense that different governments stand to lose different amounts from a trade war, with the more “powerful” governments standing to lose less. In this scenario, multilateral punishments *are* desirable because they allow for an exchange of enforcement power that bilateral punishments cannot achieve: more specifically, if punishments are multilateral, each country can offer third-party enforcement in bilateral relationships where it is “strong” in exchange for receiving third-party enforcement from other countries in bilateral relationships where it is “weak.”⁷⁴

The next point made by Maggi (1999) is that, even though some third-country punishment is in general desirable, there may be no need to make it very severe. More specifically, increasing the severity of third-country punishments beyond a certain point does not enhance cooperation, so there is no need to make these punishments “maximal.” Furthermore, the threat of third-country punishments is necessary only for certain violations, namely those by stronger countries against weaker countries, which are hard to deter with bilateral sanctions alone. The intuition is based on the result I mentioned earlier: if there are no bilateral asymmetries, there is no need for *any* third-country punishments, so it is intuitive that if there is only a limited degree of bilateral asymmetries, a limited amount of third-country punishment will be sufficient. These results seem broadly consistent with the fact that in the GATT-WTO the role of third-country punishments is more subtle and selective than the role of bilateral punishments.⁷⁵

Also Bagwell et al. (2007) focus on the design of enforcement rules in a multilateral world. In particular, this paper examines the desirability of “tradeable” retaliation rights, an idea proposed a few years ago by Mexico in the WTO. The basic idea is that small countries have a limited ability to retaliate against large countries, so allowing small countries to sell their rights of retaliation to third countries might improve the performance of the

⁷⁴ Gains from multilateral punishments can arise also from non-separabilities across bilateral relationships, which arise if there are trade-diversion effects of bilateral trade barriers. In this case, the benefits from multilateralizing punishments arise from the *aggregation* of enforcement power across trading relationships (see Maggi, 1994). A simple intuition for this effect can be gained by thinking about the effects of trade embargoes: a multilateral embargo inflicts a proportionally more severe punishment than a bilateral embargo, since the latter is partially neutralized by substitution across bilateral trade flows.

⁷⁵ There is a legitimate question as to whether third-country punishments play *any* role in the GATT-WTO. Maggi (1999) argues that they do, although in subtle ways. One form of third-country punishment for example may be the withdrawal of some “goodwill” by third countries toward the defecting country, resulting for example in a reluctance to enter new agreements with that country. Also, one should not forget that, as Thomas Schelling made clear, the effectiveness of an army sometimes must be judged by how little it is used: while it is true that full-blown multilateral retaliation has never been observed in the GATT-WTO, it is also true that there have been no cases of blatant and repeated violations of key rules, even by strong countries against vulnerable trading partners. It is reasonable to think that strong countries may have been deterred from abusing weaker partners not by the threat of bilateral retaliation, but by the implicit threat that the whole trading system may unravel as a consequence, that is, by the threat of a multilateral breakdown of cooperation.

system. Bagwell et al. consider a scenario where a violation of the agreement has already occurred and the injured country has been granted the right to retaliate, and examine two mechanisms for selling this right: a “basic” auction, in which the injured country is not allowed to bid to retire the right, and an “extended” auction, where the injured country is allowed to bid as well. Bagwell et al. find that the basic auction may “fail”, in the sense that no bids are made despite positive valuation by bidders, while the “extended” auction can never fail, and in such auction the right of retaliation is always retired. The two auction formats are then evaluated from a normative standpoint, and the ranking between them is found to depend critically on the specific normative criterion that one adopts.⁷⁶

Another interesting question of enforcement-rule design arises when countries seek to cooperate over multiple policy areas. In this case, one may ask whether there should be *issue linkage* in the enforcement of the TA, that is, whether violations in one policy area should be met with retaliation in other areas. Two papers that address this question are Ederington (2002) and Limão (2005).

Ederington (2002) considers a repeated-game model where each government can choose a tariff and a domestic policy. In this model markets are competitive, but in each country there is a localized externality which can be corrected by using the domestic policy. First, Ederington shows that domestic policies are always set at their efficient (Pigouvian) levels in the optimal self-enforcing TA, while tariffs are set above their efficient levels to accommodate self-enforcement constraints. The intuition is related to the targeting principle: since the only international externality is a TOT externality, the only reason countries are tempted to deviate from the TA is to manipulate TOT, so raising tariffs is the most efficient way to neutralize this incentive to defect.⁷⁷ The second, related result is that issue linkage need not enhance cooperation: in particular, if the punishment strategy takes the form of a permanent reversion to the one-shot Nash equilibrium (“grim-trigger” punishment), then the benefits from issue linkage are nil.

Limão (2005) considers a setting characterized not only by TOT externalities but also by cross-border pollution externalities. Governments choose tariffs and production taxes in repeated-game fashion. In this setting, Limão shows that issue linkage always allows governments to achieve a higher joint welfare relative to a non-linked agreement. This in itself may not be surprising, but a subtle question concerns how linkage affects the level of cooperation on a policy-by-policy basis. The key finding is that, if policies are independent in the governments’ objective functions, linkage promotes cooperation

⁷⁶ Bagwell et al. (2007) are agnostic about the criterion according to which the mechanisms should be evaluated, but I will not be agnostic here. In my view, the natural evaluation criterion is given by the efficiency of the equilibrium of the repeated game that the enforcement mechanism allows to achieve (see my discussion at the beginning of the section). The analysis of Bagwell et al. (2007) is silent about this criterion, simply because they do not model the repeated-game explicitly, but take it as a given that a violation has already occurred and take the amount of permissible retaliation as exogenous.

⁷⁷ This result was first derived in Ederington (2001), and then extended by Ederington (2002) to allow for both linked and non-linked agreements, as well as for different types of punishment strategies.

in one policy area at the expense of the other, because linkage effects a reallocation of enforcement power across issues; but if policies are strategic complements, then linkage can lead to more cooperation in both policy areas.⁷⁸

My next theme of discussion is the role of retaliation in the presence of asymmetric information. When governments have private information, for example about domestic political shocks, retaliation can play two distinct roles. First, retaliation can be a “punitive” tool that serves to deter blatant violations of the agreement, along the lines discussed earlier in this section. This type of retaliation is meant to be an off-equilibrium threat, so it need not be observed in equilibrium. Second, retaliation can be used for “screening” (or “truthtelling”) purposes, in the sense of inducing governments to increase protection only in states of the world where it is (politically) efficient to do so. Intuitively, retaliation imposes a cost on a government that increases protection, so if the severity of retaliation is appropriately calibrated, it may induce a government to increase protection when—and only when—the political gains from protection are high. Unlike “punitive” retaliation, this latter type of retaliation *is* meant to occur in equilibrium under some contingencies. Some scholars, for example [Schwartz and Sykes \(2002\)](#), have argued that the reciprocity rule in Article XXVIII of GATT—which provides for the “withdrawal of substantially equivalent concessions” in response to a breach of negotiated tariff bindings—can be interpreted as serving a screening function along the lines I just described. What makes this interpretation appealing is that the reciprocity rule provides only for a limited amount of retaliation, which seems consistent with the screening function, whereas the “punitive” function calls for more severe retaliation threats. I also note that the screening role of retaliation is relevant even if the agreement is perfectly enforceable, or if governments are so patient that self-enforcement constraints do not bind. Indeed, some of the papers that consider this role of retaliation, such as [Beshkar \(2010a,b\)](#) or (in a more reduced-form setting) [Maggi and Staiger \(2012, 2013\)](#), assume perfectly enforceable TAs.⁷⁹

In the context of a repeated tariff game with private information, the distinction between the punitive role and the screening role of retaliation can be understood in the following way. Suppose the importing government chooses a tariff and privately observes the value of a political-economy parameter. The agreement can be thought of

⁷⁸ Here I will also mention a paper by [Limão and Saggi \(2008\)](#) that examines whether it may be desirable to use monetary fines or bonds as part of the enforcement mechanism. Limão and Saggi show that the use of monetary fines does not help if fines must be self-enforcing and hence ultimately supported by the threat of retaliatory tariffs. On the other hand, bonds *can* enhance efficiency if they are posted with a third party prior to trading, and the third party uses the bond to compensate the injured country in case a violation is committed: intuitively, this can relax self-enforcement constraints because it indirectly injects some external enforcement into the agreement.

⁷⁹ See Section 3.1.5 for a discussion of these papers. In Maggi and Staiger’s models, retaliation is not modeled explicitly, but is captured in reduced form as a deadweight cost associated with government-to-government compensation. I also note that in these models there is no private information, but political-economy shocks are not verifiable by the court/DSB, and this form of information asymmetry has similar implications as private information for the design of incentive contracts.

as specifying a tariff *schedule* that links the tariff to the realization of the political-economy parameter. In order to be self-enforcing, the TA must discourage two types of deviations: (i) “on-schedule” deviations, whereby the government applies a tariff that is meant for another “type”; and (ii) “off-schedule” deviations, whereby the government applies a tariff level that is not meant for any “type”. In this context, one can think of the punitive role of retaliation as discouraging off-schedule deviations, while the screening role of retaliation is to discourage on-schedule deviations. A paper that focuses on these themes is [Martin and Vergote \(2008\)](#), which considers a setting where shocks are i.i.d. over time. An interesting finding of this paper is that, even if the TA can specify a “reciprocity” mechanism whereby a government’s tariff increase is met with a contemporaneous tariff response by the other government, it is always desirable for the TA to include future retaliation, provided governments are sufficiently patient. Relatedly, [Bagwell \(2009\)](#) considers a setting where political-economy shocks can be persistent over time, and shows that persistence can make enforcement more difficult, because of a “ratchet” effect whereby some government types are more reluctant to reveal themselves today for fear of being perceived as having weak retaliatory power tomorrow. For a broader and informal discussion of these themes, see also the piece by [Bagwell \(2008\)](#).

Finally, it is important to note that retaliation is not the only way to provide “truth-telling” incentives to governments. For example, [Bagwell and Staiger \(2005a\)](#) argue that this can be accomplished by imposing a dynamic constraint on the use of safeguard actions. The nature of this constraint is that, if a government invokes the escape clause today, it has to wait a certain amount of time before invoking it again. Intuitively, this can mitigate governments’ temptation to misrepresent their information and over-use the escape clause. A dynamic usage constraint of this kind is contained in the WTO’s Safeguard Agreement.⁸⁰

3.2.3. *The Enforcement Role of the DSB*

An important question concerning the enforcement of TAs is whether and how a judicial system such as the WTO’s DSB can assist with the enforcement of the TA. The answer to this question is far from obvious, since in reality TAs must be self-enforcing and international courts have no direct enforcement power. One possible answer to this question is suggested by [Maggi’s \(1999\)](#) model, which I briefly discussed in the previous section: the DSB can make multilateral enforcement possible, by disseminating information about violations of the agreement to the whole trading community. To illustrate, suppose that country A commits a violation against country B, and that in the absence of a DSB this violation would be observed only by country B. Then a potential role for the DSB is to

⁸⁰ Another potential way to induce truth-telling is introducing cross-policy linkages, that is requiring a government to make adjustments in other policies when raising trade protection. This idea plays a key role in the models by [Lee \(2007, 2011\)](#).

verify the violation and bring it to the attention of third countries, thus exposing country A to punishments by third countries.⁸¹

It is interesting to note that in the context of the WTO institution there exists another procedure that helps disseminate information about trade policies, namely the “Trade Policy Review Mechanism.” This mechanism is arguably complementary with the DSB as a way to disseminate information and improve multilateral monitoring: the former is a systematic, periodic, and wide-ranging review, whereas the DSB conducts more thorough and targeted investigations when a country (or a group of countries) files a complaint.

In practice, one important aspect of the WTO’s dispute settlement procedure is that it encourages governments to renegotiate, even following clear violations of the agreement. In Section 3.1.5, I pointed out that the renegotiation of trade policy commitments can be beneficial when the TA is an incomplete contract. However, as the repeated game literature has abundantly made clear, the renegotiation of *punishments* can have deleterious effects on cooperation, because it can undermine the credibility of punishments. This is the core of the argument in Ludema (2001), who argues that the WTO’s dispute settlement procedure can have adverse effects on cooperation if it encourages the renegotiation of punishments. Interestingly, Watson et al. (2008) make a point that goes in the opposite direction: they argue that, in the absence of a dispute settlement procedure, after a deviation countries would quickly renegotiate, thus undermining the enforcement of the agreement, whereas the presence of a dispute settlement procedure can slow down renegotiation and make it more costly, and this paradoxically can facilitate cooperation.

The above-mentioned papers raise an interesting question concerning the design of dispute settlement procedures: to what extent should they encourage renegotiation and settlement between governments? One clear point that emerges from this literature is that it is critical to distinguish between two kinds of renegotiation, namely the renegotiation of substantive *policy* rules and the renegotiation of *enforcement* rules: the former tends to be desirable, the latter tends to be harmful. But in reality the distinction between these two forms of renegotiation can be blurred, and an interesting research question would be whether and how a dispute settlement procedure can be designed to encourage one form of renegotiation and discourage the other.

4. REGIONAL TRADE AGREEMENTS

In recent years there has been a considerable amount of theoretical and empirical research on regional trade agreements (RTAs), spurred at least in part by the growing role of this

⁸¹ Another theoretical paper where the DSB facilitates the self-enforcement of TAs by changing the information structure of the game, but in a two-country setting, is Park (2011). In Park’s model, each government privately observes a noisy signal of the other government’s trade policy, and the DSB can facilitate cooperation by converting the privately observed signals into public signals. Here I also note that most of the academic research on the enforcement role of the DSB has been theoretical, but there are a few notable empirical papers on this topic, see in particular Bown (2004a), Bown (2004b), Reinhardt (2001), and Busch and Reinhardt (2002).

type of agreement in the real world. In this section I will focus on three topics: (a) the economic and political determinants of RTA formation; (b) the impact of an RTA on its members' external trade barriers and on multilateral trade liberalization; and (c) the design of rules for trade negotiations. The common theme of this section is an emphasis on the endogeneity of trade policies. I will leave aside, on the other hand, the older question of how an exogenously formed RTA affects trade flows and welfare (e.g. through trade-diversion and trade-creation effects), which has been covered extensively in [Baldwin and Venables' \(1995\)](#) handbook chapter.⁸²

4.1. Determinants of RTAs

A question of obvious positive and normative relevance concerning RTAs is: in a world where governments are motivated by political economy considerations, under what conditions are RTAs more likely to form? And in particular, is an RTA more likely to form when it is beneficial or detrimental to global welfare? This question is examined by [Grossman and Helpman \(1995b\)](#), who consider a setting similar to [Grossman and Helpman \(1995a\)](#) but with two small economies that can choose whether to form a free trade agreement (FTA). Given the presence of producer lobbies, an FTA is more likely to be adopted when it generates larger rents for the producers of both countries, a situation that Grossman and Helpman label “enhanced protection.” In a given sector, enhanced protection can occur under the following circumstances. Suppose that one country has a lower external tariff than the other, and that as a result of the FTA producers from the former country can export all of their output to the latter country without affecting local prices in the latter country. Then producers in the former country get higher rents while producers in the latter country are not affected. If the FTA generates enhanced protection in a sufficient number of sectors and in a relatively balanced way between the countries, then it will be politically viable. But since enhanced protection is more likely when the FTA causes trade diversion, the conclusion is that more trade-diverting FTAs are more likely to be adopted.

A similar question is examined by [Krishna \(1998\)](#), but in the context of an oligopolistic model with segmented markets, where governments care only about the profits of their domestic firms. In such an environment, an FTA is adopted if and only if it increases profits in both countries. This in turn is more likely to happen when the FTA leads to a bigger reduction in the market share of non-member-country firms in the member countries' markets, or in other words, when the FTA induces more trade diversion.

⁸² The theoretical research on the trade-diversion and trade-creation effects of RTAs was developed mostly before 1995, but there are some notable recent contributions, in particular [Panagariya and Krishna \(2002\)](#), which extends the classic Kemp-Wan result (which applies to customs unions) to the case of free trade agreements, and [Freund \(2000a\)](#), which compares the welfare effects of alternative (exogenous) paths of trade agreements leading to global free trade. In recent years there have been also some interesting empirical studies on the trade-diversion and trade-creation effects of RTAs: some prominent examples are [Trefler \(2004\)](#), [Romalis \(2007\)](#), [Magee \(2008\)](#), and [Clausing \(2001\)](#).

Therefore Krishna's analysis delivers a similarly pessimistic message as [Grossman and Helpman \(1995b\)](#): FTAs are more likely to be adopted when they are more trade-diverting and hence detrimental to welfare.

A counterpoint to the pessimistic message of the two papers discussed above is represented by the well-known argument (made for example by [Krugman, 1991](#)) that RTAs are more likely to form among "natural" trading partners, that is countries characterized by especially large gains from mutual trade liberalization (for example because of their geographical proximity or their comparative advantage structure). If this is the case, the argument goes, the RTAs that emerge in equilibrium will be more likely to cause trade creation than trade diversion, and hence more likely to increase welfare.

Given the seemingly contrasting theoretical arguments outlined above, the question of the welfare impact of endogenously formed RTAs is ultimately an empirical one, and indeed this question has been tackled by a number of interesting empirical papers. [Krishna \(2003\)](#) takes a structural approach to this question, estimating a general equilibrium model and using its structural parameters to examine the welfare effects of a number of hypothetical RTAs. Krishna finds that 80% of these hypothetical RTAs would be welfare-improving, but interestingly, he finds that neither geographical variables nor trade volumes are significantly correlated with the welfare gains, thus offering little support for the natural-trading-partners view of the world. [Baier and Bergstrand \(2004\)](#) find that the likelihood of an RTA is higher when countries are closer to each other and are more isolated from the rest of the world, a finding that supports the view that "natural" trading partners are more likely to form RTAs. [Baier and Bergstrand \(2007\)](#) examine the impact of RTAs on trade flows when taking into account the endogeneity of RTA formation. They find that, when this endogeneity is recognized, the trade-creation effects of RTAs appear to be much larger, and in particular, the impact of RTAs on trade flows increases fivefold relative to estimates that take RTA formation as exogenous. [Egger and Larch \(2008\)](#) find results that are consistent with those of the above-mentioned papers using a larger sample, and furthermore find that an RTA is more likely to form when there are pre-existing RTAs involving near-by countries.

Another question that has received some attention in the literature is how the likelihood of RTA formation is affected by multilateral trade liberalization. [Freund \(2000b\)](#) examines this question within a repeated-game model and finds that deeper multilateral trade liberalization leads to more RTAs, for two reasons: it increases the incentives to form an RTA and it increases the likelihood that it is self-enforcing.⁸³ At the empirical level, [Fugazza and Robert-Nicoud \(2012\)](#) find some evidence that multilateral trade liberalization increases the likelihood of subsequent RTA formation. More specifically, they find that after the Uruguay Round, the US had a higher propensity to liberalize trade on a preferential basis in goods where it had granted the deepest multilateral tariff cuts.

⁸³ See also [Ethier \(1998\)](#) for an examination of the impacts of multilateral trade liberalization on subsequent RTAs.

Finally, an interesting question is: what determines the choice between an FTA and a customs union (CU)? Empirically, most RTAs take the form of FTAs. [Facchini et al. \(2013\)](#) propose a theoretical explanation for stylized fact, based on a three-country political economy model with imperfect competition, where each country elects a representative to choose trade policies. Under a CU, since tariffs are chosen collectively by the member countries, the voters of each country strategically delegate power to a more protectionist representative. Facchini et al. show that, because of this strategic delegation effect, an FTA tends to imply higher welfare for member countries and is more likely to be politically viable than a CU.

4.2. Impacts of RTAs

In this section I will focus on two related themes: the impact of RTAs on member countries' trade barriers against outsiders ("external" trade barriers) and the impact of RTAs on multilateral trade liberalization.

The impact of RTA formation on external trade barriers has been the subject of a sizable theoretical literature. One point that emerges clearly from this literature is that the qualitative impact of RTAs on external tariffs depends crucially on whether the agreement takes the form of an FTA or of a CU. I will focus first on FTAs.

Various papers have pointed out a strong tendency of FTAs to lead to *lower* external trade barriers. This is generally known as the "tariff complementarity" effect, but it is important to point out that such effect can arise from two distinct mechanisms. The first one, highlighted by [Richardson \(1993\)](#), occurs when the FTA leads member countries to compete for tariff revenue, thus inducing them to reduce external tariffs. Interestingly, this mechanism can occur even for small countries that use tariffs only for political-economy reasons and have no TOT power. A second mechanism was pointed out by [Bagwell and Staiger \(1999b\)](#): an FTA leads member countries to import less from non-member countries, and if member countries have TOT power this reduces their incentives to manipulate TOT vis-à-vis non-members, in turn leading to lower external tariffs.⁸⁴

The tariff complementarity effect however is not the only possible effect at play in determining the impact of FTAs on external tariffs, and other effects may arise that mitigate or overturn it. [Limão \(2007\)](#) for example shows that an FTA can lead to higher external tariffs if the FTA serves also non-trade objectives (such as enhancing cooperation on labor standards or security issues). Several real-world FTAs appear to have this feature, with one country (typically the US or the EU) granting tariff preferences, and the other (typically a less developed country) making non-trade concessions. In this type of situation, the country that grants tariff preferences may be better off increasing its external

⁸⁴ Other papers that have highlighted tariff-complementarity effects in different settings are [Bond et al. \(2004\)](#), [Cadot et al. \(1999\)](#), [Bagwell and Staiger \(1997a\)](#), [Ornelas \(2005a,c\)](#), and [Saggi and Yildiz \(2010\)](#).

tariffs, because this enhances the value of the preferences and hence allows it to extract larger non-trade concessions.⁸⁵

In the case of CUs, tariff-complementarity effects may still be present, but two new forces arise that push in the opposite direction. The first one is known as the “market power” effect: if two member countries import the same good from outsiders, once the CU is formed they *jointly* have more power over TOT, and thus have a stronger incentive to raise tariffs against outsiders. The second effect is known as the “coordination” effect: if country A increases the tariff on imports of a certain good from country B, this has a positive externality on all other countries (both importers and exporters of that good), and a CU allows member countries to internalize this externality, thus leading to higher external tariffs. The first paper in the literature to highlight these two effects was [Kennan and Riezman \(1990\)](#). Note that the coordination effect can arise even in the absence of the market power effect: this is crystal-clear in a setting of “competing exporters,” where each country is the sole importer of a given good. Once all the effects are taken into account, a CU can still lead to lower external tariffs, but this is less likely than in the case of an FTA (as shown by [Bagwell and Staiger, 1999b](#)).⁸⁶

In a series of papers, Emanuel Ornelas has established a link between the literature on the impacts of RTAs and that on the determinants of RTAs, by showing that taking into account the impact of RTAs on external tariffs has important implications for the likelihood of RTA formation in the first place.

[Ornelas \(2005a\)](#) considers a model similar to [Grossman and Helpman \(1995b\)](#), but in which external tariffs are determined endogenously. Ornelas identifies a “rent destruction” effect of the FTA, which arises from the fact that the rents from external tariff protection spill over to partner countries under the FTA. The rent-destruction effect lowers the incentives of special interest groups to lobby for protection, and this creates a tendency of FTAs to induce reductions in external tariffs. Interestingly, when political-economy motivations are stronger, the drop in external tariffs is larger, thus FTAs are more conducive to multilateral trade liberalization.

The second point made by [Ornelas \(2005a\)](#) is that, since an FTA lowers the total amount of rents that trade protection can generate, this has important implications for the political viability of the FTA. Ornelas starts by considering a situation where there is no ex-ante lobbying (that is, no lobbying to influence directly the decision to join the FTA), and shows that in this case only FTAs that are sufficiently welfare-enhancing can be politically viable: intuitively, a welfare-reducing FTA cannot be attractive to the government, because it reduces both the level of welfare and the available amount of

⁸⁵ Another effect that may work against the tariff complementarity effect was highlighted by [Stoyanov \(2009\)](#): if foreign lobbying is possible, producers from FTA partners may have stronger incentives to lobby for higher external tariffs once the FTA is formed, since their gains from such external tariffs are higher under the FTA.

⁸⁶ The effects of a CU on external tariffs have been examined also by [Krugman \(1991\)](#), [Bond and Syropoulos \(1996\)](#), [Bagwell and Staiger \(1997b\)](#), and [Cadot et al. \(1999\)](#).

rents. Ornelas then allows for ex-ante lobbying, and finds that in this case a welfare-reducing FTA may in some cases be politically viable, but this is made less likely by the rent-destruction effect of the FTA: in particular, a welfare-reducing FTA can be viable only if the governments' valuation of welfare relative to contributions is neither too small nor too large. The bottom line of this paper then is that the rent-destruction effect reduces the political viability of welfare-reducing FTAs.⁸⁷

The impact of RTAs on external trade barriers has also been the subject of recent empirical work. [Estevadeordal et al. \(2008\)](#) focus on the effect of preferential trade liberalization on external tariffs in Latin America from 1990 to 2001. An appealing feature of this dataset is the wide variation in trade preferences across sectors and over time. Employing a rich set of fixed effects, these authors find that preferential tariff reduction induces faster decline in external tariffs. Furthermore, they find that this effect is present only for FTAs, not for CUs, and is stronger in sectors where the potential for trade diversion is larger.⁸⁸

Interestingly, [Limão \(2006\)](#) and [Karacaovali and Limão \(2008\)](#) find results that seemingly diverge from those of [Estevadeordal et al. \(2008\)](#). These two papers examine the impact of preferential trade liberalization by the US and the EU on multilateral trade liberalization, and find that the US and the EU liberalized less during the Uruguay Round in sectors where they had granted tariff preferences, suggesting that preferential liberalization might hinder the cause of global free trade. What can explain the difference in findings between these papers and [Estevadeordal et al. \(2008\)](#)? Theory can help us answer this question. A key difference between the two approaches is that [Limão \(2006\)](#) and [Karacaovali and Limão \(2008\)](#) focus on the US and the EU, whereas [Estevadeordal et al. \(2008\)](#) focus on developing countries. Tariffs are considerably higher in developing countries, so the potential for trade diversion is larger for these countries, and as theory suggests, this implies a stronger tariff complementarity effect. Furthermore, [Limão's \(2007\)](#) theoretical analysis suggests that preferential liberalization may hinder global free trade if RTAs are formed also for non-trade reasons, and this is more often the case for North–South RTAs than for South–South RTAs.⁸⁹

Thus far I have focused on how the formation of an RTA affects its member countries' unilateral choices of external tariffs. Next I focus on the impact of RTAs on the political viability of multilateral trade agreements.

⁸⁷ Similar results are obtained by [Ornelas \(2005b,c\)](#) in the context of an oligopolistic model similar to [Krishna \(1998\)](#).

⁸⁸ Two other papers present findings that are consistent with those of [Estevadeordal et al. \(2008\)](#): [Calvo-Pardo et al., \(2009\)](#) find that the formation of ASEAN led to a reduction in external tariffs by its member countries, and [Bohara et al. \(2004\)](#) find that the increase in preferential imports from Brazil to Argentina that followed the formation of MERCOSUR led to a decrease in Argentina's external tariffs.

⁸⁹ Related to this literature is also a recent paper by [Baldwin and Jaimovich \(2012\)](#), which focuses on the impact of RTAs on subsequent RTAs. This paper extends [Baldwin's \(1995\)](#) model of "domino" regionalism and tests its main empirical prediction, namely that the formation of an RTA has a contagion effect and increases the likelihood that further RTAs will be formed. Using a comprehensive panel of FTAs, Baldwin and Jaimovich find support for the contagion prediction.

Levy (1997) considers a model where gains from trade can arise from differences in relative factor endowments and/or from increased product variety, and takes a median-voter approach to the choice of trade policies. He shows that, if the FTA provides a country's median voter with disproportionately large gains, it may raise his or her reservation utility above the level offered by a multilateral agreement, thus undermining political support for the latter. This undermining is more likely to occur in FTAs that involve countries with similar relative factor endowments. For example, suppose Germany joins the EU. Assuming that EU countries have similar relative factor endowments, this will benefit the median voter in Germany mostly through variety gains. In the next stage, Germany considers signing a multilateral agreement. If Germany is relatively rich in capital and its median voter is an unskilled worker, the multilateral agreement is likely to damage the median voter through Stolper-Samuelson effects, without providing much additional variety gains, so he/she will block the multilateral agreement. On the other hand, if the same median voter were asked if he/she supports a multilateral agreement *before* joining the EU, then the answer may be yes, because in this case the variety gains may outweigh the adverse Stolper-Samuelson effects.

The model by Krishna (1998), already mentioned in Section 4.1, leads to a similar result, in spite of the very different structure. Recall that Krishna focuses on a Grossman-Helpman type model with oligopolistic competition. After establishing that politically viable FTAs are more likely to be trade-diverting, Krishna examines how an FTA affects the political viability of a multilateral agreement. The key finding is that the formation of an FTA may increase producers' opposition to a subsequent multilateral agreement, because the latter may reduce or eliminate the rents created by the FTA, and it can even reverse the preferences of producers, from supporting a multilateral agreement to opposing it. Thus, taken together, the papers by Levy (1997) and Krishna (1998) deliver a pessimistic message about the impact of RTAs on the political viability of multilateral agreements.

Finally, Bagwell and Staiger's (1999a) model (already discussed in Sections 2.1 and 3.1.4) also delivers pessimistic implications for the impact of RTAs on multilateral trade agreements, but for a very different reason than the models discussed above. A simple corollary of Bagwell and Staiger's analysis is that the presence of RTAs, by breaking the ability of the MFN rule to channel all international policy externalities into a single world-price externality, undermines the effectiveness of the MFN and reciprocity rules in guiding countries toward an efficient policy outcome.⁹⁰

⁹⁰ Here I will mention two other strands of literature that are quite interesting but distinct from the ones discussed in the text. The first one examines how RTAs affect the self-enforceability of a multilateral agreement. In particular, Bagwell and Staiger (1997a,b) focus on the transition period during which an RTA is being negotiated, showing that the anticipation of the RTA has an important impact on the sustainability of multilateral tariff cooperation; Saggi (2006) considers the impact of RTAs on the sustainability of a multilateral agreement in an oligopolistic setting, finding that the presence of an RTA undermines multilateral cooperation when countries are symmetric, but not necessarily when countries are asymmetric; finally, Bond et al. (2001) focus on the transition period during which trade barriers within a CU are phased out, and examine how the deepening of intra-CU trade liberalization affects the sustainability

4.3. Rules for Trade Negotiations

In this section I return to my emphasis on the design of rules, but this time I focus on rules that constrain trade negotiations rather than the policy choices of individual governments. A number of papers have been written on the question of what rules (if any) should regulate trade negotiations, but this literature can be hard to tame, as the modeling approach and the exact nature of the question seem to shift from paper to paper. In what follows I propose a simple conceptual framework that might be useful to organize our thinking and define more clearly the relevant questions.

Consider the following scenario: imagine that, at some ex-ante stage, all countries involved in the trading system bargain over the rules that regulate future trade negotiations. Also assume that at the ex-ante stage countries cannot write a complete agreement, and in particular they cannot specify actual trade policies, but only negotiation rules (otherwise there would be no reason to set negotiation rules in the first place). One possible reason why negotiation rules might be desirable is that governments may have incentives to sign RTAs that exert negative externalities on non-member countries, in which case the resulting outcome may be globally inefficient.

In the scenario I just described, one can make a distinction between two types of rules constraining trade negotiations: (1) Rules that constrain the nature of the agreement itself. An example of this type of rule is one that prohibits RTAs, or equivalently, a rule requiring that any agreement be approved by all countries. Another example is a rule that prohibits CUs but not FTAs. I refer to these rules (for lack of a better expression) as *rules-to-make-rules*; (2) Rules that constrain the policy content of the agreements. Examples include reciprocity-type rules, which require that tariff reductions be balanced across countries, and non-discrimination rules, such as the MFN rule. These rules do not constrain the type of agreements that countries sign, but do constrain the policies (or policy changes) that countries can agree upon. I refer to these as *policy-content* rules.⁹¹

Note that the GATT-WTO currently does not impose rules-to-make-rules, but only policy-content rules, and more specifically: (i) GATT Article I imposes the MFN rule,

of multilateral agreements. The other line of research I want to mention is the work by [Martin et al. \(2008, 2012\)](#), who explore the two-way relationship between RTAs and military conflicts. At the theoretical level, they argue that RTAs increase the opportunity cost of conflict, thereby reducing the likelihood of war, and conversely, countries with a higher likelihood of conflict are more likely to sign RTAs as a way to promote peace. Furthermore, they argue that this logic does not apply to multilateral trade agreements, because multilateral trade openness decreases bilateral dependence from trade with any given country and hence the cost of a bilateral conflict. At the empirical level, they find support for these predictions using a large dataset of military conflicts during the 1950–2000 period.

⁹¹ In the scenario outlined above I am assuming that at the ex-ante stage countries can only specify rules constraining future negotiations, not rules constraining individual government policies. In reality, of course, even from the very beginning the GATT imposed both types of rules. I am separating the stages at which the two different types of rules are designed for conceptual simplicity. Also, the distinction between policy-content rules and rules-to-make-rules can be subtle but should be conceptually clear. For example, note that imposing the MFN rule does not logically imply prohibiting FTAs, because an FTA in principle need not violate MFN (an FTA eliminates tariffs between its members, but this does not prevent a member from respecting the MFN rule, which it can do by eliminating the relevant tariffs vis-à-vis non-members).

so any negotiated policy changes must be extended in a non-discriminatory way to all member countries; (ii) GATT Article XXIV allows an exception to the MFN rule for the case of an RTA that eliminates substantially all trade barriers among its members, but requires the RTA member countries not to raise their external tariffs above pre-RTA levels;⁹² and (iii) GATT requires that negotiations adhere to the principle of reciprocity, although as discussed earlier, this is an informal principle rather than a strict rule.

It is important, however, to consider also the possibility of rules-to-make-rules, both from a positive perspective (why are there no such rules in GATT-WTO?) and from a normative one (should such rules be imposed?). In particular, the simplest such rule—a ban on RTAs—has a special theoretical interest. Indeed, an interesting question can be posed here: if there are no frictions in multilateral bargaining (as assumed in most of the existing models), why would a ban on RTAs *not* be optimal? If governments are forced to choose trade policies at a multilateral bargaining table, why would this not achieve efficiency?⁹³ I will return to this question at the end of this section, after surveying the recent research in this area.

If one adopts the conceptual framework outlined above, one needs to take a stand on what objective function the rules are supposed to maximize. There are two possible questions that can be asked, each of which is interesting in its own right:

1. What rules maximize the likelihood of reaching global free trade? This is essentially the question raised by Bhagwati (1993): are RTAs “building blocks” or “stumbling blocks” on the path to global free trade?
2. What rules maximize the joint payoff of all governments? This is a more positive question and can lead to different answers than the previous one, if one allows for political motivations in the governments’ objectives.

Ideally, a full answer to either question requires comparing the *equilibrium configuration* of trade agreements (or their path, if one takes a dynamic approach) under alternative negotiation rules, of course allowing for RTAs as well as multilateral agreements. Models that focus on the impact of exogenously formed RTAs on multilateral agreements, such as those that I surveyed in the previous section, evidently can only shed partial light on this question.

4.3.1. Policy-Content Rules

As I mentioned earlier, the key policy-content rules that regulate trade negotiations in the GATT-WTO system are the MFN rule and—at a more informal level—the reciprocity principle. In Section 3.1.4, I discussed papers by Bagwell and Staiger (1999a) and

⁹² Exceptions to MFN are also allowed in a number of extraordinary circumstances, such as national-security or health hazards (GATT Article III).

⁹³ Of course it is possible that such a rule is not strictly needed, if global efficiency can be achieved also in the absence of any rule on trade negotiations, and in this case a ban on RTAs will only be *weakly* optimal. Indeed, this is the case in some of the models I will survey below.

Ossa (2011) that examine the implications of these rules in the context of multilateral trade negotiations, but these papers do not consider endogenous RTAs.

A paper that does consider the impact of these rules on the endogenous formation of RTAs is Bagwell and Staiger (2005b). This paper considers a two-stage scenario where, in the first stage, countries can sign a multilateral agreement, and in the second stage, pairs of countries can sign bilateral agreements. The first point of the paper is that, if bilateral negotiations are left unrestricted, a problem of “bilateral opportunism” is likely to arise: after multilateral trade concessions have been exchanged, there is an incentive for a pair of countries to take a further step and liberalize trade bilaterally, but this will erode the value of the concessions that the excluded country had obtained in the initial multilateral negotiation, and this in turn makes countries more reluctant to make multilateral trade concessions in the first place. The second point of the paper is that the bilateral opportunism problem described above may be solved if trade negotiations are disciplined by the MFN rule in conjunction with a reciprocity rule. To gain intuition for this result, suppose there are only two goods. Then the MFN rule ensures the existence of a single relative world price (as explained in Section 2.1.2), and reciprocity ensures that this world price is effectively fixed by the initial multilateral agreement; and given that the world price is preserved in subsequent bilateral negotiations, the welfare of countries that do not participate in a bilateral negotiation is preserved as well.

In a related paper, Bagwell and Staiger (2010b) consider a setting in which countries can sequentially sign bilateral agreements in the presence of the MFN rule. Two inefficiencies tend to arise in this setting: the first one is due to the bilateral opportunism problem highlighted above; the second is that, since later negotiating partners can free-ride on the MFN concessions that a country makes to early negotiating partners, a country may be induced to offer too little in the early negotiations (“foot-dragging”). Bagwell and Staiger then argue that these inefficiencies can be removed if two additional rules are imposed, along with MFN: the reciprocity rule and a “non-violation” rule along the lines of GATT Article XXIII.1b. These rules together act as a device to “secure” the concessions received by a country in early negotiations and protect them from potential free riding and bilateral opportunism in the future.⁹⁴

A different argument for the desirability of the MFN rule is proposed by McCalman (2002). This contribution is notable because it is a rare example of a model that explicitly introduces bargaining frictions in multilateral trade negotiations. In this model, a large country negotiates with N small countries over tariffs and transfers, and each small country has private information about its gains from an agreement. McCalman compares two scenarios, one where the large country can make different offers to different countries, and one where the MFN rule constrains the large country to make the same offer to all countries. The large country of course is worse off under the MFN rule, since the

⁹⁴ This is a good juncture to mention a paper by Ludema and Mayda (2009), which examines empirically the free-rider problem associated with the MFN rule, finding that this problem is indeed of first-order empirical importance.

latter constrains its choice, but global efficiency may be higher under MFN, and this is more likely when N is larger. The reason lies in the fact that bargaining is inefficient, due to private information. If bargaining were frictionless, unconstrained bargaining would always lead to efficiency, and the MFN rule would always be a bad idea. But in the presence of private information, unconstrained bargaining is inefficient, because the large country is not able to appropriate the entire surplus from the negotiations, and as a consequence it is possible that imposing MFN increases the welfare of the N countries more than it reduces the welfare of the large country.⁹⁵

4.3.2. Rules-to-Make-Rules

The simplest rule-to-make-rules, and the one that has received the most attention in the literature, is a rule that prohibits RTAs. Would the world be more efficient if all agreements had to be multilateral in nature? In this section I discuss a number of recent papers that speak to this question.

One of the first papers to examine the desirability of a rule banning RTAs is McLaren (2002), who focuses on the implications of irreversible investments for the formation of trade agreements. McLaren considers a world with three countries: in the first period, individuals choose in what sectors to allocate their resources; in the second period, there is a coalition-formation game among governments, which can yield an FTA, a multilateral free trade agreement, or no agreement at all. Negotiations are costly, and this cost is higher for a multilateral negotiation. This model may have multiple equilibria, including an equilibrium where global free trade arises (and the allocation is efficient), and equilibria where two of the countries form an FTA (and the allocation is inefficient). The latter type of equilibrium is based on a self-fulfilling prophecy. If individuals expect an FTA between countries A and B, they will make investment decisions that make these countries more specialized relative to each other, thus increasing the gains from trade between them. At the same time, countries A and B will become less specialized relative to the outside country, thus decreasing the gains from trade between them and the outside country. As a consequence, ex-post countries A and B will have a strong incentive to sign an FTA, while the benefits from a multilateral agreement will be small. Ex-ante, an FTA equilibrium of this kind may be inefficient if the allocation distortions associated with the FTA outweigh the savings in negotiation costs. Intuitively, then, there is a region of parameters in which an FTA equilibrium exists and is inefficient, and therefore a ban on FTAs is strictly desirable.

⁹⁵ Another policy-content rule imposed by the GATT-WTO, as I mentioned in the previous section, is contained in GATT Article XXIV. Mrazova et al. (2013) examine the implications of Article XXIV as it applies to CUs, requiring that the common external tariff of the CU must not exceed the average of its members' pre-CU tariffs. Mrazova et al. argue that Article XXIV increases the probability that free trade emerges in equilibrium, but when free trade does not arise in equilibrium, the constraints imposed by this rule may lead to a reduction in world welfare.

Goyal and Joshi (2006) adopt a network-formation approach to study the formation of FTAs, using a particular notion of “stability” to determine the configurations of FTAs that can arise in equilibrium. Focusing on a setting with a homogenous good and symmetric countries, Goyal and Joshi find that the complete FTA network, which yields global free trade, is a stable network, thus suggesting that FTAs are “building blocks” toward global free trade. Another network-theoretic model of FTA formation is Furusawa and Konishi (2007), which differs from Goyal and Joshi (2006) in that it allows for differentiated goods and a richer pattern of asymmetries between countries. Consistently with Goyal and Joshi, this paper finds that when countries are symmetric, the complete FTA network (global free trade) is stable. However, if goods are highly substitutable, there may also be other stable networks that do not yield global free trade. And if countries are asymmetric, the complete FTA network may not be stable. Furusawa and Konishi conclude that FTAs may under some conditions be “stumbling blocks” toward global free trade.⁹⁶

Before moving on, I will make a general observation about the network-theoretic approach to the analysis of RTA formation, of which I just discussed two examples. In these network models, there is no meaningful distinction between a sequence of RTAs that leads to global free trade and a multilateral agreement. Indeed, multilateral agreements *per se* are not considered at all, so these models, though capable of generating interesting insights, are arguably not well equipped to evaluate the desirability of rules-to-make-rules, such as a ban on RTAs. The models that I discuss next, on the other hand, do allow for multilateral agreements as well as RTAs, and hence are better equipped to examine this question.

Seidmann (2009) examines a three-country bargaining model in which countries can negotiate FTAs, CUs, and multilateral agreements. Countries are allowed to continue negotiating after reaching an agreement, and for this reason, an RTA can be used by its member countries to improve their bargaining position for subsequent trade negotiations (“strategic positioning” effect). Two important features of the model are that international transfers are available and that global free trade maximizes the three countries’ joint surplus. Seidmann studies the equilibrium configuration of agreements, highlighting for example conditions under which a hub-and-spoke structure emerges. Regarding the question of whether RTAs are building blocks or stumbling blocks toward global free trade, in Seidmann’s model global free trade may or may not be reached if RTAs are

⁹⁶ This is a good juncture to mention a paper by Yi (1996), who takes a coalition-formation approach to study the formation of CUs. Yi compares two possible games: a “unanimous regionalism” game, where a CU forms if and only if all potential members agree to form the CU; and an “open regionalism” game, where each country chooses an “address,” and then all the countries that have chosen the same address must be part of the same CU. Yi shows that the grand CU (global free trade) is an equilibrium of the open-regionalism game, but typically is not an equilibrium of the unanimous-regionalism game, and interprets this finding as suggesting that an “open regionalism” rule is desirable. However it is not clear how to interpret Yi’s notion of open regionalism in a way that has a meaningful counterpart in the real world: Yi’s open-regionalism game implicitly assumes that a country is not free to *leave* a CU (if this were the case then we would be in the unanimous-regionalism game), which seems like a far-fetched idea.

feasible, while global free trade is always reached if RTAs are not feasible, thus a ban on RTAs is always weakly desirable.

Saggi and Yildiz (2010) also consider a three-country bargaining model where governments can negotiate RTAs as well as multilateral agreements. The negotiation game is as follows: governments simultaneously name FTA partners, and an FTA is formed if the announcements agree; if a government names both of the other countries, this is interpreted as a proposal for a multilateral agreement, and if all governments propose a multilateral agreement, it is implemented. A key feature of the model is that international transfers are not available. Saggi and Yildiz examine the implications of a rule banning FTAs, by comparing a game where FTAs are allowed with one where they are not. Focusing on coalition-proof Nash equilibria, they show that when countries are symmetric global free trade is the only stable equilibrium, whether or not FTAs are allowed. But when countries are asymmetric it may happen that global free trade is a stable equilibrium *only if* FTAs are allowed, so a ban on FTAs can make global free trade less likely. To understand this result intuitively, recall that there are no international transfers in the model, so global free trade is not the only Pareto-efficient outcome. Indeed, a ban on FTAs may lead away from global free trade, but *not* away from Pareto-efficiency. If FTAs are banned, it is possible that in equilibrium two countries agree to liberalize trade while the third does not: this outcome is skewed in favor of the country that free rides, but is Pareto efficient. Thus, Saggi and Yildiz's result should be interpreted as suggesting only that FTAs may be needed to achieve global free trade, not that FTAs may be needed to achieve global efficiency.

Saggi et al. (2013) build on Saggi and Yildiz (2010) by considering the case of CUs. The main finding of this paper is that, in contrast with the case of FTAs, in the case of CUs a “stumbling block” scenario is possible, in the sense that the freedom to pursue CUs may prevent the attainment of global free trade. Interestingly, the reason for this difference in results is not that a CU has a more harmful impact on outsiders than an FTA, but rather that it implies a stronger incentive for insiders to deny access to outsiders. Taken together, the two papers just discussed suggest that under some conditions it might be desirable to ban CUs but not FTAs; but this is only suggestive, because in order to make this point rigorously one would need to consider a model where CUs and FTAs are *both* options available to governments, which is not the case in either of the above-mentioned papers.

Aghion et al. (2007) also examine the building-block vs stumbling-block question within a three-country model where countries can negotiate FTAs as well as multilateral agreements, but with some key differences relative to Saggi and Yildiz (2010). In particular, Aghion et al. assume that there is a leading country that chooses whether to engage in sequential bilateral bargains or in a single multilateral bargain, and allow for political economy motivations in the government objectives. In addition, they assume that international transfers are available. Aghion et al. define payoffs to be “grand-coalition superadditive” if the payoff of the grand coalition is larger than the payoff of all countries

combined in alternative coalition structures. This property is satisfied for example if free trade is Pareto-efficient and each government maximizes national welfare. A key result of the paper is that, if payoffs are grand-coalition superadditive, then the leading country may prefer sequential bargaining or multilateral bargaining, depending on the nature of coalition externalities, but in either case global free trade must emerge in equilibrium.

Aghion et al. then examine environments where grand-coalition superadditivity fails, which can happen when political-economy motivations are strong. In this case, it is possible that global free trade is attained only if FTAs are permitted (a building-block scenario), and it is also possible that global free trade is attained only if FTAs are banned (a stumbling-block scenario). Note that, as in [Saggi and Yildiz \(2010\)](#), it may happen that a ban on FTAs leads away from global free trade, but not that a ban on FTAs leads away from Pareto-efficiency. However, unlike Saggi-Yildiz, this is not due to the absence of international transfers, but rather to the possibility that governments may not maximize welfare; indeed, a ban on FTAs can lead away from free trade only if governments do not maximize welfare.

A related point is made by [Ornelas \(2008\)](#): if governments do not maximize welfare, it is possible that FTAs have the effect of bringing the world closer to global free trade relative to a multilateral agreement, and hence a ban on RTAs may lead the world away from global free trade. However, this possibility arises in [Ornelas \(2008\)](#) for very different reasons than in the papers mentioned above: here the reasons are that FTAs lead member countries to lower their external tariffs, and that there is a tendency for FTAs to emerge in equilibrium when they are trade-creating rather than trade-diverting.

A common message suggested by the papers discussed above is that, if rules are designed ex-ante when governments are in “constitution-writing” mode and seek to maximize global welfare, but ex-post government objectives may diverge from welfare, then a ban on RTAs may be harmful, because RTAs may lead the world closer to global free trade than a multilateral agreement.

I conclude this section by returning to the question I posed earlier: why has the GATT-WTO *not* banned RTAs? As discussed above, the possible answers suggested by the literature are two: (i) because rules are designed ex-ante when governments are in constitution-writing mode and maximize welfare, whereas ex-post government objectives may diverge from welfare; or (ii) because efficiency might also be achieved by means of other rules, such as MFN and reciprocity. A third, and conceptually simpler, reason why banning RTAs might not be a good idea is the presence of important frictions in multilateral bargaining. This consideration is arguably of first-order empirical importance, but has received little attention in the formal literature.⁹⁷ If multilateral bargaining is less efficient than bilateral bargaining, for example because of the complexity of negotiations when a large number of countries is involved, then allowing RTAs may be strictly

⁹⁷ One notable exception is [McLaren \(2002\)](#), but he models multilateral bargaining frictions in a very reduced-form way, through a parameter that captures the extra cost of multilateral negotiations.

desirable on efficiency grounds. But a more complete understanding of this issue would require introducing bargaining frictions explicitly in our models and examining how they depend on the set of countries and on the set of issues involved in the bargain.

5. CONCLUSION

Coming back to the pessimistic statement made by [Krugman \(1997\)](#) (see beginning of Section 2), one is tempted to ask: have the last 15 years of research proved Krugman wrong? This is a matter of debate, but I think that, at a minimum, the literature has demonstrated that the logic of economics can go a long way toward explaining the purpose and design of trade agreements. As I have argued in this chapter, more progress is needed along several dimensions, but we have made significant advances toward understanding the motivations that drive countries to sign trade agreements and the reasons why trade agreements are designed the way they are, and we are now in a better position to evaluate possible reforms of existing rules from a normative standpoint. Most of the research has been at the theoretical level, but in the last few years there has been an acceleration in empirical research, spurred by the availability of new and better data sets.

What's next for the economics of trade agreements? At a broad level, I think that this research area is ready—both in terms of theoretical tools and data availability—to follow a path that other research areas in international economics have fruitfully taken, namely a tighter integration between theoretical and empirical analysis. Some of the most important questions on the table require counterfactual analysis, and this in turn calls for structural and/or calibration approaches. Some recent papers that I discussed in this chapter have moved in this direction, but we are still at the beginnings.

In terms of substantive questions, I will point to a few directions that seem promising to me. One question that we still know little about is the empirical importance of “New Trade” and domestic-commitment motives for trade agreements. Another important set of open questions concerns multilateral trade negotiations in the presence of bargaining frictions: why has the Doha round failed? Is it because there are no mutual gains left on the table, or because of bargaining frictions due to the large number of countries and issues on the table? If bargaining frictions are part of the problem, can bargaining protocols be designed in a smarter way to facilitate more efficient outcomes?

If history is of any guidance, the release of new important datasets can trigger new waves of empirical and theoretical research. This will hopefully be the case for the recent release by the WTO of an unprecedented dataset that includes extremely detailed information about the bargaining that took place during GATT negotiation rounds. This dataset may help answer new questions, such as those related to the nature of bargaining frictions and the importance of bargaining protocols, as well as old questions, such as the extent to which the MFN rule generates free-rider problems and whether bilateral trade agreements are building blocks or stumbling blocks toward global free trade.

REFERENCES

- Aghion, P., Antràs, P., Helpman, E., 2007. Negotiating free trade. *Journal of International Economics* 73 (1), 1–30.
- Amador, M., Bagwell, K., 2013. The theory of optimal delegation with an application to tariff caps. *Econometrica* 81 (4), 1541–1599.
- Amador, M., Bagwell, K., 2012. Tariff revenue and tariff caps. *American Economic Review Papers and Proceedings* 102 (3), 459–465.
- Antràs, P., Staiger, R.W., 2012a. Offshoring and the role of trade agreements. *American Economic Review* 102 (7), 3140–3183.
- Antràs, P., Staiger, R.W., 2012b. Trade agreements and the nature of international price determination. *American Economic Review Papers and Proceedings* 102 (3), 470–476.
- Bacchetta, M., Piermartini, R., 2011. The Value of Bindings. WTO Staff Working Paper, ERS-D-2011-13.
- Bagwell, K., 2008. Remedies in the WTO: an economic perspective. In: Janow, M.E., Donaldson, V.J., Yanovich, A. (Eds.), *The WTO Governance, Dispute Settlement & Developing Countries*. Juris Publishing, pp. 733–770 (Chapter 40).
- Bagwell, K., 2009. Self-Enforcing Trade Agreements and Private Information. NBER Working Paper 14812.
- Bagwell, K., Staiger, R.W., 1990. A theory of managed trade. *The American Economic Review* 80, 779–795.
- Bagwell, K., Staiger, R.W., 1997a. Multilateral tariff cooperation during the formation of free trade areas. *International Economic Review* 38, 291–319.
- Bagwell, K., Staiger, R.W., 1997b. Multilateral tariff cooperation during the formation of customs unions. *Journal of International Economics* 42, 91–123.
- Bagwell, K., Staiger, R.W., 1999a. An economic theory of GATT. *American Economic Review* 89 (1), 215–248.
- Bagwell, K., Staiger, R.W., 1999b. Regionalism and multilateral tariff cooperation. In: Piggott, J., Woodland, A. (Eds.), *International Trade Policy and the Pacific Rim*. MacMillan, London, pp. 157–185.
- Bagwell, K., Staiger, R.W., 2001. Domestic policies, national sovereignty and international economic institutions. *Quarterly Journal of Economics* 116 (2), 519–562.
- Bagwell, K., Staiger, R.W., 2002. *The Economics of the World Trading System*. MIT Press, Cambridge, MA.
- Bagwell, K., Staiger, R.W., 2003. Protection over the business cycle. *Advances in Economic Analysis and Policy* 3 (1), 1–43.
- Bagwell, K., Staiger, R.W., 2005a. Enforcement, private political pressure, and the GATT/WTO escape clause. *Journal of Legal Studies* 34 (2), 471–513.
- Bagwell, K., Staiger, R.W., 2005b. Multilateral trade negotiations, bilateral opportunism and the rules of GATT/WTO. *Journal of International Economics* 67, 268–294.
- Bagwell, K., Staiger, R.W., 2006. Will international rules on subsidies disrupt the world trading system? *American Economic Review* 96 (3), 877–895.
- Bagwell, K., Staiger, R.W., 2009. *Delocation and Trade Agreements in Imperfectly Competitive Markets*. Mimeo.
- Bagwell, K., Staiger, R.W., 2010a. The WTO: theory and practice. *Annual Review of Economics* 2, 223–256.
- Bagwell, K., Staiger, R.W., 2010b. Backward stealing and forward manipulation in the WTO. *Journal of International Economics* 82 (1), 49–62.
- Bagwell, K., Staiger, R.W., 2011. What do trade negotiators negotiate about? Empirical evidence from the world trade organization. *American Economic Review* 101 (4), 1238–1273.
- Bagwell, K., Staiger, R.W., 2012a. Profit shifting and trade agreements in imperfectly competitive markets. *International Economic Review* 53 (4), 1067–1104.
- Bagwell, K., Staiger, R.W., 2012b. The economics of trade agreements in the linear delocation model. *Journal of International Economics* 88 (1), 32–46.
- Bagwell, K., Mavroidis, P.C., Staiger, R.W., 2007. Auctioning counter-measures in the WTO. *Journal of International Economics* 73 (2), 309–332.
- Baier, S.L., Bergstrand, J.H., 2004. Economic determinants of free trade agreements. *Journal of International Economics* 64 (1), 29–63.

- Baier, S.L., Bergstrand, J.H., 2007. Do free trade agreements actually increase members' international trade? *Journal of International Economics* 71 (1), 72–95.
- Baldwin, R.E., 1987. Politically realistic objective functions and trade policy. *Economics Letters* 24 (1), 287–290.
- Baldwin, R.E., 1995. A domino theory of regionalism. In: Baldwin, R., Haaparanta, P., Kiander, J. (Eds.), *Expanding membership of the European Union*. Cambridge University Press, pp. 25–53.
- Baldwin, R.E., Jaimovich, D., 2012. Are free trade agreements contagious? *Journal of International Economics* 88 (1), 1–16.
- Baldwin, R.E., Venables, A.J., 1995. Regional economic integration. In: Grossman, G., Rogoff, K. (Eds.), *Handbook of International Economics*, vol. 3. North-Holland, Amsterdam, pp. 1597–1644.
- Battigalli, P., Maggi, G., 2002. Rigidity, discretion and the costs of writing contracts. *American Economic Review* 92 (4), 798–817.
- Battigalli, P., Maggi, G., 2003. *International Agreements on Product Standards: An Incomplete-Contracting Theory*. NBER Working Paper 9533.
- Bajona, C., Chu, T., 2010. Reforming state owned enterprises in China: effects of WTO accession. *Review of Economic Dynamics* 13 (4), 800–823.
- Bajona, C., Ederington, J., 2012. Domestic Policies, Hidden Protection and the GATT/WTO, Mimeo.
- Beshkar, M., 2010a. Optimal remedies in international trade agreements. *European Economic Review* 54 (3), 455–466.
- Beshkar, M., 2010b. Trade skirmishes and safeguards: a theory of the WTO dispute settlement process. *Journal of International Economics* 82 (1), 35–48.
- Beshkar, M., 2011. Arbitration and Renegotiation in Trade Agreements, Mimeo.
- Beshkar, M., Bond, E.W., 2012. Cap and Escape in Trade Agreements, Mimeo.
- Beshkar, M., Bond, E.W., Rho, Y., 2011. Tariff Binding and Overhang: Theory and Evidence, Mimeo.
- Bhagwati, J., 1993. Regionalism and multilateralism: an overview. In: de Melo, J., Panagariya, A. (Eds.), *New Dimensions in Regional Integration*. Cambridge University Press, pp. 22–51.
- Bohara, A.K., Gawande, K., Sanguinetti, P., 2004. Trade diversion and declining tariffs: evidence from Mercosur. *Journal of International Economics* 64, 65–88.
- Bond, E.W., Park, J.H., 2004. Gradualism in trade agreements with asymmetric countries. *Review of Economic Studies* 69, 379–406.
- Bond, E.W., Syropoulos, C., 1996. The size of trading blocks: market power and world welfare effects. *Journal of International Economics* 40, 412–437.
- Bond, E.W., Syropoulos, C., Winters, L.A., 2001. Deepening of regional integration and multilateral trade agreements. *Journal of International Economics* 53, 335–362.
- Bond, E.W., Riezman, R., Syropoulos, C., 2004. A strategic and welfare theoretic analysis of free trade areas. *Journal of International Economics* 64, 1–27.
- Bown, C., 2004a. On the economic success of GATT/WTO dispute settlement. *Review of Economics and Statistics* 86, 811–823.
- Bown, C., 2004b. Trade disputes and the implementation of protection under the GATT: an empirical assessment. *Journal of International Economics* 62, 263–294.
- Bown, C., Crowley, M.A., 2006. Policy externalities: how US antidumping affects Japanese exports to the EU. *European Journal of Political Economy* 22, 696–714.
- Bown, C.P., Crowley, M.A., 2013a. Self-enforcing trade agreements: evidence from time-varying trade policy. *American Economic Review* 103 (2), 1071–1090.
- Bown, C.P., Crowley, M.A., 2013b. Import protection, business cycles, and exchange rates: evidence from the great recession. *Journal of International Economics* 90 (1), 50–64.
- Brander, J., Spencer, B., 1984. Tariff protection and imperfect competition. In: Kierzkowski, H. (Ed.), *Monopolistic Competition and International Trade*. Clarendon Press, Oxford.
- Brander, J., Spencer, B., 1985. Export subsidies and market share rivalry. *Journal of International Economics* 18, 83–100.
- Broda, C., Weinstein, D., 2006. Globalization and the gains from variety. *Quarterly Journal of Economics* 121 (2), 541–585.
- Broda, C., Limão, N., Weinstein, D.E., 2008. Optimal tariffs and market power: the evidence. *American Economic Review* 98 (5), 2032–2065.

- Brou, D., Ruta, M., 2009. A Commitment Theory of Subsidy Agreements, Mimeo.
- Busch, M., Reinhardt, E., 2002. Testing international trade law: empirical studies of GATT/WTO dispute settlement. In: Kennedy, D.L.M., Southwick, J.D. (Eds.), *The Political Economy of International Trade Law: Essays in Honor of Robert E. Hudec*. Cambridge University Press, pp. 457–481.
- Cadot, O., de Melo, J., Olarreaga, M., 1999. Regional integration and lobbying for tariffs against non-members. *International Economic Review* 40, 635–657.
- Calvo-Pardo, H., Freund, C., Ornelas, E., 2009. The ASEAN free trade agreement: impact on trade flows and external trade barriers. In: Barro, R., Lee, J. (Eds.), *Costs and Benefits of Regional Economic Integration*. Oxford University Press.
- Chisik, R., 2003. Gradualism in free trade agreements: a theoretical justification. *Journal of International Economics* 59 (2), 367–397.
- Clausing, K., 2001. Trade creation and trade diversion in the Canada–United States Free Trade Agreement. *Canadian Journal of Economics* 34, 678–696.
- Conconi, P., Perroni, C., 2009. Do credible domestic institutions promote credible international agreements? *Journal of International Economics* 79, 160–170.
- Conconi, P., Sahuguet, N., 2009. Policymakers' horizon and the sustainability of international cooperation. *Journal of Public Economics* 93, 549–558.
- Conconi, P., Facchini, G., Zanardi, M., 2012. Fast-track authority and international trade negotiations. *American Economic Journal: Economic Policy* 4, 146–189.
- Costinot, A., 2008. A comparative institutional analysis of agreements on product standards. *Journal of International Economics* 75 (1), 197–213.
- Dekle, R., Eaton, J., Kortum, S., 2007. Unbalanced trade. *American Economic Review Papers and Proceedings* 97 (2), 351–355.
- DeRemer, D., 2011. *The Evolution of International Subsidy Rules*, Mimeo.
- Devereux, M.B., 1997. Growth, specialization and trade liberalization. *International Economic Review* 38, 565–585.
- Dutt, P., Mihov, I., Van Zandt, T., 2011. Does WTO Matter for the Extensive and the Intensive Margins of Trade? Mimeo.
- Ederington, J., 2001. International coordination of trade and domestic policies. *American Economic Review* 91, 1580–1593.
- Ederington, J., 2002. Trade and domestic policy linkage in international agreements. *International Economic Review* 43, 1347–1367.
- Egger, P., Larch, M., 2008. Interdependent preferential trade agreement memberships: an empirical analysis. *Journal of International Economics* 76, 384–399.
- Eicher, T.S., Henn, C., 2011. In search of WTO trade effects: preferential trade agreements promote trade strongly, but unevenly. *Journal of International Economics*, 83 (2), 137–153.
- Estevadeordal, A., Freund, C., Ornelas, E., 2008. Does regionalism affect trade liberalization toward non-members? *The Quarterly Journal of Economics* 123 (4), 1531–1575.
- Ethier, W.J., 1998. Regionalism in a multilateral world. *Journal of Political Economy* 106, 1214–1245.
- Facchini, G., Silva, P., Willmann, G., 2013. The customs union issue: why do we observe so few of them? *Journal of International Economics* 90 (1), 136–147.
- Freund, C., 2000a. Different paths to free trade: the gains from regionalism. *Quarterly Journal of Economics* 115, 1317–1341.
- Freund, C., 2000b. Multilateralism and the endogenous formation of preferential trade agreements. *Journal of International Economics* 52 (2), 359–376.
- Furusawa, T., 1999. The Negotiation of sustainable tariffs. *Journal of International Economics* 48, 321–346.
- Furusawa, T., Konishi, H., 2007. Free trade networks. *Journal of International Economics* 72, 310–335.
- Furusawa, T., Lai, E.L.-C., 1999. Adjustment costs and gradual trade liberalization. *Journal of International Economics* 49, 333–361.
- Gawande, K., Krishna, P., Robbins, M.J., 2006. Foreign lobbying and US trade policy. *The Review of Economics and Statistics* 88 (3), 563–571.
- Goldberg, P., Maggi, G., 1999. Protection for sale: an empirical investigation. *American Economic Review* 89 (5), 1135–1155.
- Goyal, S., Joshi, S., 2006. Bilateralism and free trade. *International Economic Review* 47, 749–778.

- Grossman, G., Helpman, E., 1994. Protection for sale. *American Economic Review* 84 (4), 833–850.
- Grossman, G., Helpman, E., 1995a. Trade wars and trade talks. *Journal of Political Economy* 103 (4), 675–708.
- Grossman, G., Helpman, E., 1995b. The politics of free-trade agreements. *American Economic Review* 85, 667–690.
- Handley, K., 2012. Exporting Under Trade Policy Uncertainty: Theory and Evidence. Mimeo.
- Handley, K., Limão, N., 2012. Trade and Investment under Policy Uncertainty: Theory and Firm Evidence, NBER WP 17790.
- Handley, K., Limão, N., 2013. Policy Uncertainty, Trade and Welfare: Theory and Evidence for China, Mimeo.
- Horn, H., 2006. National treatment in the GATT. *American Economic Review* 96 (1), 394–404.
- Horn, H., Maggi, G., Staiger, R.W., 2010. Trade agreements as endogenously incomplete contracts. *American Economic Review* 100 (1), 394–419.
- Jackson, J.H., 1997. The WTO dispute settlement understanding—misunderstandings on the nature of legal obligation. *American Journal of International Law* 91 (1), 60–64.
- Jackson, J.H., 2006. Sovereignty, the WTO, and changing fundamentals of international law. Cambridge University Press.
- Johnson, H.G., 1953. Optimum tariffs and retaliation. *Review of Economic Studies* 21 (2), 142–153.
- Karacaovali, B., Limão, N., 2008. The clash of liberalizations: preferential vs. multilateral trade liberalization in the European Union. *Journal of International Economics* 74 (2), 299–327.
- Kennan, J., Riezman, R., 1990. Optimal tariff equilibria with customs unions. *Canadian Journal of Economics* 23, 70–83.
- Kreinin, M.E., 1961. Effect of tariff changes on the prices and volume of imports. *American Economic Review* 51, 310–324.
- Krishna, P., 1998. Regionalism and multilateralism: a political economy approach. *Quarterly Journal of Economics* 113, 227–251.
- Krishna, P., 2003. Are regional trading partners natural? *Journal of Political Economy* 111, 202–226.
- Krishna, P., Mitra, D., 2005. Reciprocated unilateralism in trade policy. *Journal of International Economics* 65 (2), 461–487.
- Krugman, P., 1980. Scale economies, product differentiation, and the pattern of trade. *The American Economic Review* 70 (5), 950–959.
- Krugman, P., 1991. Is bilateralism bad? In: Helpman, E., Razin, A. (Eds.), *International Trade and Trade Policy*. MIT Press, pp. 9–23.
- Krugman, P., 1997. What should trade negotiators negotiate about? *Journal of Economic Literature* 35, 113–120.
- Lapan, H.E., 1988. The optimal tariff, production lags, and time consistency. *American Economic Review* 78 (3), 395–401.
- Lee, G.M., 2007. Trade agreements with domestic policies as disguised protection. *Journal of International Economics* 71 (1), 241–259.
- Lee, G.M., 2011. Optimal International Agreement and Treatment of Domestic Subsidy, Mimeo.
- Levy, P.I., 1997. A political-economic analysis of free-trade agreements. *American Economic Review* 87, 506–519.
- Limão, N., 2005. Trade policy, cross-border externalities and lobbies: do linked agreements enforce more cooperative outcomes? *Journal of International Economics* 67 (1), 175–199.
- Limão, N., 2006. Preferential trade agreements as stumbling blocks for multilateral trade liberalization: evidence for the US. *American Economic Review* 96, 896–914.
- Limão, N., 2007. Are preferential trade agreements with non-trade objectives a stumbling block for multilateral liberalization? *Review of Economic Studies* 74, 821–855.
- Limão, N., Maggi, G., 2013. Uncertainty and Trade Agreements, Mimeo.
- Limão, N., Saggi, K., 2008. Tariff retaliation versus financial compensation in the enforcement of international trade agreements. *Journal of International Economics* 76 (1), 48–60.
- Limão, N., Tovar, P., 2011. Policy choice: theory and evidence from commitment via international trade agreements. *Journal of International Economics* 85 (2), 186–205.
- Liu, X., 2009. GATT/WTO promotes trade strongly: sample selection and model specification. *Review of International Economics* 17 (3), 428–446.

- Liu, X., Ornelas, E., 2012. Free Trade Areas and the Consolidation of Democracies, Mimeo.
- Ludema, R.D., 2001. Optimal international trade agreements and dispute settlement procedures. *European Journal of Political Economy* 17 (2), 355–376.
- Ludema, R.D., Mayda, A.M., 2009. Do countries free ride on MFN? *Journal of International Economics* 77, 135–150.
- Ludema, R.D., Mayda, A.M., 2010. Do Terms-of-Trade Effects Matter for Trade Agreements? Evidence from WTO Countries, Mimeo.
- Ludema R.D., Mayda, A.M., Mishra, P., 2010. Protection for Free: The Political Economy of US Tariff Suspensions, Mimeo.
- Magee, C., 2008. New measures of trade creation and trade diversion. *Journal of International Economics* 75, 340–362.
- Maggi, G., 1994. Essays on Trade Policy and International Institutions under Incomplete Information. Ph.D. Dissertation, Stanford University.
- Maggi, G., 1999. The role of multilateral institutions in international trade cooperation. *American Economic Review* 89 (1), 190–214.
- Maggi, G., Rodriguez-Clare, A., 1998. The value of trade agreements in the presence of political pressures. *Journal of Political Economy* 106 (3), 574–601.
- Maggi, G., Rodriguez-Clare, A., 2007. A political-economy theory of trade agreements. *American Economic Review* 97 (4), 1374–1406.
- Maggi, G., Staiger, R.W., 2011. The role of dispute settlement procedures in international trade agreements. *Quarterly Journal of Economics* 126, 475–515.
- Maggi, G., Staiger, R.W., 2012. Optimal Design of Trade Agreements in the Presence of Renegotiation, Mimeo.
- Maggi, G., Staiger, R.W., 2013. Trade Disputes and Settlement, Mimeo.
- Martin, A., Vergote, W., 2008. On the role of retaliation in trade agreements. *Journal of International Economics* 76 (1), 61–77.
- Martin, P., Mayer, T., Thoenig, M., 2008. Make trade not war? *Review of Economic Studies* 75 (3), 865–900.
- Martin, P., Mayer, T., Thoenig, M., 2012. The geography of conflicts and regional trade agreements. *American Economic Journal: Macroeconomics* 4 (4), 1–35.
- Mayer, W., 1981. Theoretical considerations on negotiated tariff adjustments. *Oxford Economic Papers* 33 (1), 135–153.
- McCalman, P., 2002. Multilateral trade negotiations and the most favored nation clause. *Journal of International Economics* 57 (1), 151–176.
- McLaren, J., 1997. Size, sunk costs, and judge Bowker's objection to free trade. *American Economic Review* 87 (3), 400–420.
- McLaren, J., 2002. A theory of insidious regionalism. *Quarterly Journal of Economics* 117, 571–608.
- Milner, H., Rosendorff, P., 1996. Trade negotiations, information and domestic politics: the role of domestic groups. *Economics and Politics* 8 (2), 145–189.
- Milner, H., Rosendorff, P., 2001. The optimal design of international institutions: uncertainty and escape. *International Organization* 55 (4), 829–857.
- Mitra, D., 2002. Endogenous political organization and the value of trade agreements. *Journal of International Economics* 57 (2), 473–485.
- Mrazova, M., 2011. Trade Agreements When Profits Matter, Mimeo.
- Mrazova, M., Vines, D., Zissimos, B., 2013. Is the WTO Article XXIV bad? *Journal of International Economics* 89 (1), 216–232.
- Ornelas, E., 2005a. Rent destruction and the political viability of free trade agreements. *Quarterly Journal of Economics* 120, 1475–1506.
- Ornelas, E., 2005b. Endogenous free trade agreements and the multilateral trading system. *Journal of International Economics* 67, 471–497.
- Ornelas, E., 2005c. Trade creating free trade areas and the undermining of multilateralism. *European Economic Review* 49, 1717–1735.
- Ornelas, E., 2008. Feasible multilateralism and the effects of regionalism. *Journal of International Economics* 74, 202–224.

- Ossa, R., 2011. A new trade theory of GATT/WTO negotiations. *Journal of Political Economy* 119 (1), 122–152.
- Ossa, R., 2012. Profits in the new trade approach to trade negotiations. *American Economic Review, Papers and Proceedings* 102 (3), 466–469.
- Ossa, R., 2013. Trade Wars and Trade Talks with Data, Mimeo.
- Panagariya, A., Krishna, P., 2002. On necessarily welfare-enhancing free trade areas. *Journal of International Economics* 57, 353–367.
- Park, J.H., 2000. International trade agreements between countries of asymmetric size. *Journal of International Economics* 50 (2), 473–495.
- Park, J.H., 2011. Enforcing international trade agreements with imperfect private monitoring. *Review of Economic Studies* 78 (3), 1102–1134.
- Pauwelyn, J., 2008. *Optimal Protection of International Law: Navigating between European Absolutism and American Volunteerism*. Cambridge University Press.
- Reinhardt, E., 2001. Adjudication without enforcement in GATT disputes. *Journal of Conflict Resolution* 45 (2), 174–195.
- Richardson, M., 1993. Endogenous protection and trade diversion. *Journal of International Economics* 34 (3–4), 309–324.
- Romalis, J., 2007. NAFTA's and CUSFTA's impact on international trade. *Review of Economics and Statistics* 89, 416–435.
- Rose, A., 2004a. Do we really know that the WTO increases trade? *American Economic Review* 94, 98–114.
- Rose, A., 2004b. Do WTO members have more liberal trade policy? *Journal of International Economics* 63 (2), 209–235.
- Rose, A., 2013. *The March of an Economic Idea? Protectionism Isn't Counter-Cyclic (Anymore)*, Mimeo.
- Saggi, K., 2006. Preferential trade agreements and multilateral tariff cooperation. *International Economic Review* 47 (1), 29–57.
- Saggi, K., Sara, N., 2008. National treatment at the WTO: the roles of product and country heterogeneity. *International Economic Review* 49, 1367–1396.
- Saggi, K., Yildiz, H.M., 2010. Bilateralism, multilateralism, and the quest for global free trade. *Journal of International Economics* 81, 26–37.
- Saggi, K., Woodland, A., Yildiz, H.M., 2013. On the relationship between preferential and multilateral trade liberalization: the case of customs unions. *American Economic Journal: Microeconomics* 5 (1), 63–99.
- Schwartz, W.F., Sykes, A.O., 2002. The economic structure of renegotiation and dispute resolution in the WTO-GATT system. *Journal of Legal Studies* 31, 179–204.
- Segal, I.R., Whinston, M.D., 2002. The Mirrlees approach to mechanism design with renegotiation. *Econometrica* 70 (1), 1–45.
- Seidmann, D., 2009. Preferential trading arrangements as strategic positioning. *Journal of International Economics* 79, 143–159.
- Sovey, A., 2012. *States Held Hostage: Political Hold-Up Problems in International Politics*, Mimeo.
- Staiger, R.W., 1995a. International rules and institutions for trade policy. In: Grossman, G.M., Rogoff, K. (Eds.), *Handbook of International Economics*, vol. 3. North-Holland, Amsterdam.
- Staiger, R.W., 1995b. A theory of gradual trade liberalization. In: Deardorff, A., Levinsohn, J., Stern, R. (Eds.), *New Directions in Trade Theory*. University of Michigan Press.
- Staiger, R.W., Sykes, A.O., 2011. International Trade, National Treatment and Domestic Regulation. *Journal of Legal Studies* 40 (1), 149–203.
- Staiger, R.W., Tabellini, G., 1987. Discretionary trade policy and excessive protection. *American Economic Review* 77 (5), 823–837.
- Staiger, R.W., Tabellini, G., 1999. Do GATT rules help governments make domestic commitments? *Economics and Politics* 11 (2), 109–144.
- Stoyanov, A., 2009. Trade policy of a free trade agreement in the presence of foreign lobbying. *Journal of International Economics* 77 (1), 37–49.
- Subramanian, A., Wei, S.J., 2007. The WTO promotes trade, strongly but unevenly. *Journal of International Economics* 72, 151–175.

- Tang, M.K., Wei, S.J., 2010. The value of making commitments externally: evidence from WTO accessions. *Journal of International Economics* 78 (2), 216–229.
- Tomz, M., Goldstein, J.L., Rivers, D., 2007. Do we really know that the WTO increases trade? Comment. *American Economic Review* 97, 2005–2018.
- Tornell, A., 1991. Time inconsistency of protectionist programs. *Quarterly Journal of Economics* 106, 963–974.
- Tovar, P., 2011. Lobbying costs and trade policy. *Journal of International Economics* 83 (2), 126–136.
- Trefler, D., 2004. The long and short of the Canada–US free trade agreement. *American Economic Review* 94, 870–895.
- Venables, A.J., 1985. Trade and trade policy with imperfect competition: the case of identical products and free entry. *Journal of International Economics* 19, 1–20.
- Venables, A.J., 1987. Trade and trade policy with differentiated products: a Chamberlinian–Ricardian model. *Economic Journal* 97, 700–717.
- Watson, J., Klimenko, M.M., Ramey, G., 2008. Recurrent trade agreements and the value of external enforcement. *Journal of International Economics* 74 (2), 475–499.
- Whalley, J., 1998. Why do countries seek regional trade agreements? In: Frankel, J. (Ed.), *The Regionalization of the World Economy*. University of Chicago Press.
- Winters, A.L., Chang, W., 2000. Regional integration and the prices of imports: an empirical investigation. *Journal of International Economics* 51, 363–377.
- Winters, A.L., Chang, W., 2002. How regional blocks affect excluded countries: the price effects of Mercusor. *American Economic Review* 92, 889–904.
- Yi, S.S., 1996. Endogenous formation of customs unions under imperfect competition: open regionalism is good. *Journal of International Economics* 41, 153–177.