CHAPTER 8

The Design of Trade Agreements

K. Bagwell[*,‡], **R.W. Staiger**[†,‡]
[*]Stanford University, Stanford, CA, United States
[†]Dartmouth College, Hanover, NH, United States
[‡]NBER, Cambridge, MA, United States

Contents

Handbook of Commercial Policy, Volume 1A
ISSN 2214-3122, http://dx.doi.org/10.1016/bs.hescop.2016.04.005

Abstract

What does economics have to say about the design of international trade agreements? We review a literature on this question, providing detailed coverage on three key design features of the GATT/WTO: reciprocity, nondiscrimination as embodied in the MFN principle, and tariff bindings and binding "overhang." Each of these features is central to the design of the GATT/WTO and we argue that an economic perspective can go a long way toward revealing a consistent logic to the inclusion of these design features in trade agreements.

Keywords

Trade agreements, GATT, WTO, Reciprocity, MFN, Nondiscrimination, Tariff bindings, Binding overhang

JEL Classification Codes:

F02, F13

1. INTRODUCTION

What does economics have to say about the design of international trade agreements? In this chapter, we review a literature related to this question. We provide detailed coverage on three key design features of the General Agreement on Tariffs and Trade (GATT) and its successor the World Trade Organization (WTO).[a] The three design features on which we focus are reciprocity, nondiscrimination as embodied in the most-favored-nation (MFN) principle, and tariff bindings and binding "overhang." Each of these features is central to the design of GATT/WTO.

We adopt the view that design reflects purpose. We thus discuss the purpose of a trade agreement as a tool to set the stage for our discussion of design.[b] That is, we seek first to catalog the "problems" that a trade agreement may "solve" in the various formal models of trade agreements, where the problems lead to inefficiencies whose solutions can then generate increases in joint surplus that make a mutually beneficial trade agreement possible. With the problems identified and the inefficiencies characterized, we are then better able to assess whether according to these formal models the trade agreement is well designed to facilitate mutual gains for member governments.

We now provide the narrative for the organization of our chapter. At the broadest level, trade agreements could be designed following either of two standard traditions in economics for addressing inefficiencies (see Hoekman and Kostecki, 1995, pp. 59–60).

[a] Other design features are briefly mentioned in Section 6 and are treated in detail in other chapters of this Handbook.

[b] Our discussion of the purpose of trade agreements is thus narrow in scope. The broad literature on the purpose of trade agreements is considered by Grossman (2016).

A "top down" approach would create a supranational authority that sets trade policies for each member country. A "bottom up" approach would entail Coasean bargaining among governments, and a critical element to ensure efficiency would be the existence of secure property rights over the objects on which bargaining was to occur. In practice international trade agreements are typically designed according to the second approach: each government has property rights over its own policy instruments, but further rules may be needed to secure the relevant property rights (eg, to "market access"); governments negotiate first over the rules (multilaterally); and they then bargain over actual policies within rules (bilaterally, typically, though not always). The agreement must be self-enforcing.

From here, we can imagine how negotiators would start by agreeing that they should use policy instruments that are transparent, so that they know what is agreed upon and can monitor compliance. As we discuss below, if the diagnosis of the problem is a terms-of-trade problem at any rate, they would also focus on reciprocal liberalization. They would then need to confront the problem of third-party externalities, leading to consideration of nondiscrimination. Next, a basic question would be whether they are negotiating exact tariffs or bounds. At this point, they might have various "what-if" questions, concerning upward flexibility to shocks, opportunistic use of nonborder measures, the rules for settling disputes when unanticipated issues arise, export policy instruments, and the like. We refer to other Handbook chapters for full treatments of these important topics, and only briefly touch on them here. As indicated, our focus here is on evaluating and interpreting the design features of reciprocity, MFN and tariff bindings and overhang.

A natural question at this point is whether there is really any consistent logic to the design of trade agreements at all. Do the design features of these agreements appear to be sufficiently "purposeful" and deliberate to support the view that they can be usefully analyzed from an economic perspective? We argue in this chapter that the answer is yes.[c] To develop this argument, we present formal models that speak to the purpose and essential design of the GATT/WTO, and we also relate the implications of these models to legal and historical writings that concern the purpose and design of trade agreements.

The rest of the chapter proceeds as follows. In Section 2 we present a sequence of models and diagnose the problem that a trade agreement might solve. Building on our findings from Section 2, Section 3 then covers reciprocity, Section 4 covers MFN, and Section 5 covers bindings and overhang. Finally, Section 6 touches briefly on other important design features and concludes.

2. DIAGNOSIS OF THE PROBLEM

In this section, we present a sequence of two-country models and diagnose the problem that a trade agreement might solve. We describe each model in detail, and organize our discussion around three questions. First, is the Nash equilibrium inefficient? In effect, this

[c] See Krugman (1997) for an early dissenting view.

question asks whether there is a problem for governments that a trade agreement might solve. Second, if the Nash equilibrium is inefficient, then is reciprocal trade liberalization a necessary design feature for a mutually advantageous trade agreement among governments? By asking this question, we address an initial and basic design feature while postponing consideration of other design features until later sections. Finally, if governments were not motivated by the terms-of-trade implications of their trade policies, would the resulting noncooperative tariffs be efficient? This question acknowledges the central role played by the terms-of-trade externality in the trade-agreement literature and provides a means of categorizing models based upon whether this externality alone explains the purpose of a trade agreement.

Before proceeding, we pause to remark on the value of characterizing the purpose of a trade agreement. Why should we bother to identify the problem that a trade agreement solves? For our purposes in this chapter, the reason is this: by identifying the problem that a trade agreement may solve, we are better able to assess whether the trade agreement is well designed for facilitating mutual gains for member governments. Our discussion here thus serves as a foundation for later sections, where we consider trade-agreement design in greater detail.

We conclude the section by approaching issues of purpose and design from a different angle. Given the prominence played by the terms-of-trade externality in the formal literature on trade agreements, it is important to ask whether there is meaningful contact between the problems emphasized by this literature and those emphasized by economists and others who were directly involved in GATT's design. Drawing on an early GATT document, we describe some suggestive evidence consistent with a central role for terms-of-trade externalities.

2.1 Competitive General-Equilibrium Model of Trade Agreements

We begin with the standard perfectly competitive general-equilibrium model of trade agreements. This model provides a general framework in which to understand the basic terms-of-trade-driven Prisoners' Dilemma problem that a trade agreement may solve.

2.1.1 The Model

We consider a standard general-equilibrium model of trade between two countries in two goods. Each good is a normal good in consumption, and production is determined in perfectly competitive markets under conditions of increasing opportunity costs. The home country imports good x and exports good y, while the foreign country imports good y and exports good x. The local relative price facing domestic producers and consumers is denoted as $p \equiv p_x/p_y$, and similarly the local relative price in the foreign country is represented as $p^* \equiv p_x^*/p_y^*$, where here and throughout we use an "*" to denote foreign-country variables. The government of each country has available an ad valorem import tariff, and we assume that governments set tariffs at nonprohibitive levels.

Let $t > -1$ denote the ad valorem import tariff selection of the home government, with $t^* > -1$ representing the corresponding selection by the foreign government. It is convenient to define $\tau \equiv 1 + t$ and $\tau^* \equiv 1 + t^*$. We further define the relative world (ie, offshore or untaxed) price as $p^w \equiv p_x^*/p_y$. The world price is thus the relative price of the foreign country's export good to its import good on the "world" market and as such is the foreign country's "terms of trade." The home country's terms of trade is then $1/p^w$. With these definitions in place, we may now observe the following relationships: $p = \tau p^w \equiv p(\tau, p^w)$ and $p^* = p^w/\tau^* \equiv p^*(\tau^*, p^w)$.

Once a country's local price and terms of trade are determined, its production, consumption and tariff revenue are implied. In each country, the production of good i, where $i = x, y$, is determined by the point on the concave production possibilities frontier at which the marginal rate of transformation between x and y equals the local relative price. We may thus represent the domestic and foreign production functions as $Q_i(p)$ and $Q_i^*(p^*)$, respectively. Consumption in each country is determined by the local and world prices: $C_i(p, p^w)$ and $C_i^*(p^*, p^w)$. Intuitively, the local price determines the relative price faced by consumers and also the level and distribution of factor income, while tariff revenue is distributed lump sum to consumers and can be expressed as a function of the local and world prices.[d] We may now define import demand and export supply for the home country as $M(p, p^w) \equiv C_x(p, p^w) - Q_x(p)$ and $E(p, p^w) \equiv Q_y(p) - C_y(p, p^w)$, respectively. Similarly, we may represent the foreign country's import demand and export supply functions, respectively, as $M^*(p^*, p^w) \equiv C_y^*(p^*, p^w) - Q_y^*(p^*)$ and $E^*(p^*, p^w) \equiv Q_x^*(p^*) - C_x^*(p^*, p^w)$.

For any local and world prices, each country must satisfy budget or "trade balance" constraints. For the home country, the trade balance constraint is

$$p^w M(p, p^w) = E(p, p^w). \tag{1}$$

The foreign-country trade balance is similarly represented as

$$M^*(p^*, p^w) = p^w E^*(p^*, p^w). \tag{2}$$

We may think of these relationships as constraints that are embedded in the construction of the import demand and export supply functions.

The final ingredient in our model is a market-clearing requirement. Given τ and τ^*, we require that the world price is set so as to achieve market clearing in good x :

$$M(p(\tau, p^w), p^w) = E^*(p^*(\tau^*, p^w), p^w). \tag{3}$$

For given τ and τ^*, we notice that (3) describes one equation in a single unknown variable, p^w. Let $\tilde{p}^w(\tau, \tau^*)$ denote the market-clearing world price that satisfies (3). It is now straightforward to confirm that conditions (1), (2) and (3) ensure that market clearing is

[d] For further details, see Bagwell and Staiger (2002, pp. 14–15).

achieved in good y as well. In this general fashion, for any specification of τ and τ^*, we may determine the equilibrium world price and the associated local prices. All equilibrium quantities (production, consumption, tariff revenue, imports, exports) are then implied in turn.

The only assumption that we place upon the general-equilibrium model is that prices respond to tariffs in the "regular" way.[e] Specifically, we make the following assumptions:

$$\begin{aligned} \frac{dp(\tau,\tilde{p}^w(\tau,\tau^*))}{d\tau} > 0 > \frac{\partial \tilde{p}^w(\tau,\tau^*)}{\partial \tau} \\ \frac{dp^*(\tau^*,\tilde{p}^w(\tau,\tau^*))}{d\tau^*} < 0 < \frac{\partial \tilde{p}^w(\tau,\tau^*)}{\partial \tau^*}. \end{aligned} \tag{4}$$

In short, when a country's tariff is increased, the relative price of its import good increases in the local market and falls in the world market. The latter (world price) effect amounts to an assumption that the country is "large" and can exercise monopsony power by raising its tariff. It also means that a country can improve its terms of trade—and thus cause a deterioration in its trading partner's terms of trade—by raising its tariff.

With the basic general-equilibrium model of trade described, we now consider the preferences of governments. In the *traditional approach* to trade agreements, governments are assumed to maximize national economic welfare. This is the approach explored by Dixit (1987), Johnson (1953-1954), Kennan and Riezman (1988), and Mayer (1981), for example. Following Bagwell and Staiger (1999, 2002), we adopt a *political-economic approach* which includes the traditional approach as a special case but allows as well that governments may have political or distributional concerns. We thus represent the preferences of a government as a general function of its country's local price and terms of trade. Formally, the home and foreign government preferences, respectively, are represented as $W(p,\tilde{p}^w)$ and $W^*(p^*,\tilde{p}^w)$, where all prices are evaluated henceforth at market-clearing levels. We note that each government's welfare is ultimately a function of the underlying tariff choices, since under market clearing we have that $p = p(\tau,\tilde{p}^w)$, $p^* = p^*(\tau^*,\tilde{p}^w)$ and $\tilde{p}^w = \tilde{p}^w(\tau,\tau^*)$. It is nevertheless convenient to represent government welfare functions in terms of local and world prices, as we are thereby able to identify the channel through which one government's tariff selection imposes an externality on the welfare of the other government.

Following Bagwell and Staiger (1999, 2002), the only assumption we make on government preferences is as follows: a government benefits from a terms-of-trade improvement, when the local price in its country is held fixed. Formally, our assumption may be stated as follows:

[e] Put differently, we assume that the model does not exhibit the Metzler or Lerner paradox.

$$W_{\tilde{p}^w}(p,\tilde{p}^w) < 0 < W^*_{\tilde{p}^w}(p^*,\tilde{p}^w). \tag{5}$$

Notice that no assumption is made as to the manner in which welfare varies with the local price, and so a wide variety of political and distributional motivations can be accommodated. It is important to reflect on the meaning of the preference assumption captured in (5). Imagine that the home government raises its tariff, τ, and that the foreign government cuts its tariff, τ^*. Under (4), this adjustment in tariffs results in a lower value for $\tilde{p}^w$ and thus an improvement in the home country's terms of trade. Further, the tariff changes can be made in such a way as to maintain the local price, p, in the home country.[f] Such a change in tariffs does not alter the local price faced by producers and consumers in the home country; instead, it amounts to a transfer of tariff revenue from the foreign country to the home country. The meaning of the preference assumption in (5) is simply that a government benefits from being the recipient of such a transfer.

As Bagwell and Staiger (1999, 2002) discuss, assumption (5) is satisfied under the traditional approach where governments maximize national economic welfare. The assumption is also satisfied in the leading models that adopt the political-economic approach, including the lobbying models of Grossman and Helpman (1994, 1995) and the median-voter model of Mayer (1984).[g] As discussed below, the model presented here does not include, however, models in which governments face a "commitment problem" and have time-inconsistent preferences, since such models allow that government preferences may change over time.

2.1.2 Prisoners' Dilemma

We show next that the general-equilibrium model generates a terms-of-trade-driven Prisoners' Dilemma problem between governments. To begin, let us consider a simultaneous-move game in which governments select import tariffs. Assuming that a unique, interior Nash equilibrium exists, we represent the non-cooperative or *Nash* tariffs as a pair, (τ^N, τ^{*N}), satisfying the following first-order conditions:[h]

[f] Under (4), the increase in τ raises p and the decrease in τ^* further lowers $\tilde{p}^w$ and thus decreases p.

[g] As noted earlier, work in the traditional approach includes Dixit (1987), Johnson (1953-1954), Kennan and Riezman (1988), and Mayer (1981). Work in which government preferences correspond to the preferences of the median voter includes Dhingra (2014), Dutt and Mitra (2002), and Mayer (1984), while work in which government preferences reflect the interests of lobbies includes Baldwin (1987), Brock and Magee (1978), Caves (1976), Feenstra and Bhagwati (1982), Findlay and Wellisz (1982), Grossman and Helpman (1994, 1995), Hillman (1982), and Olson (1965). The preference assumption in (5) is satisfied in all of these approaches.

[h] As Dixit (1987) establishes for a setting in which governments maximize national economic welfare, Nash equilibria with autarky also exist. The equilibrium upon which we focus here, by contrast, has positive trade volume and is in this sense "interior."

$$\frac{dW(p,\widetilde{p}^w)}{d\tau} = W_p\frac{dp}{d\tau} + W_{\widetilde{p}^w}\frac{\partial\widetilde{p}^w}{\partial\tau} = 0$$
$$\frac{dW^*(p^*,\widetilde{p}^w)}{d\tau^*} = W^*_{p^*}\frac{dp^*}{d\tau^*} + W^*_{\widetilde{p}^w}\frac{\partial\widetilde{p}^w}{\partial\tau^*} = 0. \tag{6}$$

The first condition in (6) defines the "optimal tariff" or the best-response tariff for the home government, while the second condition defines the analogous tariff for the foreign government.[i] Using (4) and (5), we may immediately observe that $W_p < 0$ when the home government selects its optimal tariff. Intuitively, the home government would welcome the lower local price and corresponding greater trade volume that a tariff reduction would induce; yet, the home government refrains from unilaterally lowering its tariff from the optimal level, since a lower tariff would worsen its terms of trade. Similarly, for the foreign government, (4) and (5) imply that $W^*_{p^*} > 0$ at the optimal tariff, where a higher value for p^* would result from a lower foreign tariff.

A trade agreement is an agreement between governments. To understand the rationale for a trade agreement, we thus consider whether a trade agreement could generate greater government welfare than governments enjoy under non-cooperative tariff setting (ie, in the Nash equilibrium). We are thus motivated to consider Pareto efficient tariff pairs, where efficiency is measured relative to government welfare. An *efficient* pair of tariffs is defined by a tangency condition for government indifference curves:

$$\frac{d\tau}{d\tau^*}\Big|_{dW=0} = \frac{d\tau}{d\tau^*}\Big|_{dW^*=0}. \tag{7}$$

This tangency condition that defines an efficient pair of tariffs can be rewritten as

$$\frac{[\tau W_p + W_{\widetilde{p}^w}]\dfrac{\partial\widetilde{p}^w}{\partial\tau^*}}{W_p\dfrac{dp}{d\tau} + W_{\widetilde{p}^w}\dfrac{\partial\widetilde{p}^w}{\partial\tau}} = \frac{W^*_{p^*}\dfrac{dp^*}{d\tau^*} + W^*_{\widetilde{p}^w}\dfrac{\partial\widetilde{p}^w}{\partial\tau^*}}{\left[\dfrac{1}{\tau^*}W^*_{p^*} + W^*_{\widetilde{p}^w}\right]\dfrac{\partial\widetilde{p}^w}{\partial\tau}}. \tag{8}$$

Notice that, in assuming that a trade agreement may deliver an efficient tariff pair, we are putting enforcement issues to the side for now.

Under the traditional approach, the optimal import tariff for each government is positive, and so $\tau^N > 1$ and $\tau^{*N} > 1$. This finding dates back to Torrens (1844) and Mill (1844) and was formalized by Johnson (1953-1954). It implies that free trade is not the optimal unilateral tariff for a government that maximizes national economic welfare and presides over a large country. As Mayer (1981) showed, the efficiency frontier under the traditional approach is defined by the locus $\tau = 1/\tau^*$. This tariff locus ensures that

[i] Throughout, we assume that all second-order conditions are globally satisfied and that all partial derivatives of W and W^* are continuous and finite.

local prices are equalized across countries (ie, $p = p^*$) and includes global free trade ($\tau = \tau^* = 1$) as a special case. As we raise the home-country tariff and lower the foreign-country tariff while moving along the locus of efficient tariffs, the home (foreign)-country experiences a terms of trade gain (loss). Hence, the world price and thus the distribution of income across countries varies along the efficiency frontier.

The traditional approach thus indicates that governments of large countries face a Prisoners' Dilemma problem. The Nash tariffs are too high, and governments could achieve a Pareto gain by forming an agreement in which tariffs are reduced in an appropriate reciprocal manner. If governments are symmetric, then Nash and efficient tariffs lie along the 45-degree line. In this case, governments can achieve efficiency and both gain relative to the Nash equilibrium by moving to global free trade. As Johnson (1953-1954), Kennan and Riezman (1988), and Mayer (1981) argue, however, if countries are sufficiently asymmetric, then a large country may prefer the Nash equilibrium to global free trade. Thus, reciprocal tariff reductions that lead to an efficient outcome do not always generate Pareto gains relative to the Nash equilibrium.

Adopting the more general political-economic approach, Bagwell and Staiger (1999, 2002) establish the following three findings. First, and as may be easily verified, the Nash equilibrium tariffs defined by (6) do not satisfy (8) and are thus inefficient. This finding is not entirely surprising, since a higher tariff from one country imposes a negative international externality in the form of a terms-of-trade loss on the other country. The second finding is that, starting at the Nash equilibrium tariffs (τ^N, τ^{*N}), governments can mutually gain from moving to a new pair of tariffs (τ, τ^*) only if the new tariffs entail *reciprocal trade liberalization*: $\tau < \tau^N$ and $\tau^* < \tau^{*N}$. A general form of reciprocity is thus necessary if governments are to achieve mutual gains from a trade agreement.

The second finding follows easily once it is established that a government experiences a strict welfare reduction along its best-response curve as its trading partner's tariff is increased. To establish this point, we focus on the home country and define its best-response function, $\tau = \tau^R(\tau^*)$, as the solution to the first equation in (6). Following Bagwell and Staiger (1999, 2002), we may now use (6) to find that

$$\begin{aligned}\frac{dW(p,\tilde{p}^w)}{d\tau^*}\Big|_{\tau=\tau^R(\tau^*)} &= \left[(W_p\tau + W_{\tilde{p}^w})\frac{\partial\tilde{p}^w}{\partial\tau^*}\right]\Big|_{\tau=\tau^R(\tau^*)} \\ &= \left[(1-\tau\lambda)W_{\tilde{p}^w}\frac{\partial\tilde{p}^w}{\partial\tau^*}\right]\Big|_{\tau=\tau^R(\tau^*)} \\ &< 0,\end{aligned} \tag{9}$$

where $\lambda \equiv \dfrac{\partial\tilde{p}^w}{\partial\tau}\Big/\dfrac{dp}{d\tau} < 0$ by (4). To complete the argument, we suppose that a trade agreement generates mutual gains and specifies a tariff pair (τ, τ^*) for which $\tau^* > \tau^{*N}$. Starting

at this tariff pair, we may reposition the home tariff to the home best-response level, $\tau^R(\tau^*)$, and then reduce the foreign tariff from τ^* to τ^{*N} while adjusting the home tariff along the home best-response curve. It then follows from (9) that the home government experiences strictly higher welfare at the Nash tariffs than at the tariff pair specified by the agreement, which contradicts the supposition that the agreement generate mutual gains. A similar argument applies when the trade agreement specifies a pair (τ, τ^*) where $\tau > \tau^N$.

The third finding concerns the reason that the Nash tariffs are inefficient. To identify the source of the problem, Bagwell and Staiger (1999, 2002) define *politically optimal* tariffs as the tariff pair, (τ^{PO}, τ^{*PO}), that satisfies

$$W_p = 0 = W^*_{p^*}. \tag{10}$$

We may understand the politically optimal tariffs to be the tariffs that governments would choose if, hypothetically, they were not motivated by the terms-of-trade implications of their tariff selections (ie, if, hypothetically, they acted as if $W_{\tilde{p}^w} = 0 = W^*_{\tilde{p}^w}$). The third finding, which follows easily from (8) and (10), is that the politically optimal tariffs are efficient. Bagwell and Staiger (1999, 2002) interpret this finding as establishing that the terms-of-trade externality is the sole rationale for a trade agreement, even when governments have political-economic preferences. They note further that the politically optimal tariffs correspond to global free trade when governments maximize national economic welfare.

The three findings are captured in Fig. 1 (adopted from Bagwell and Staiger, 1999). As this figure illustrates, Nash tariffs are inefficient and too high. Further, reciprocal tariff liberalization (ie, $\tau < \tau^N$ and $\tau^* < \tau^{*N}$) is necessary but not sufficient for mutual gains

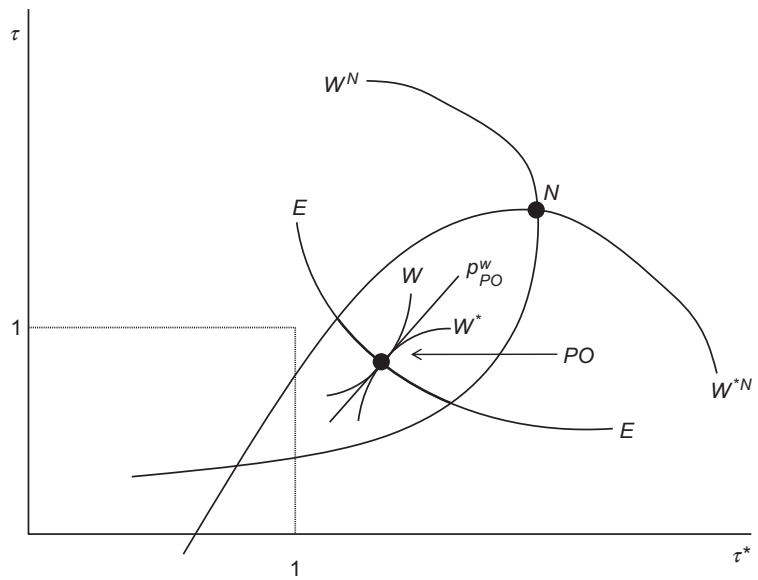

Fig. 1 The problem for a trade agreement to solve.

for governments relative to the Nash equilibrium. Finally, politically optimal tariffs are efficient. As depicted, when the home and foreign countries are sufficiently symmetric, the politically optimal tariffs reside on the contract curve (ie, they are efficient and generate greater-than-Nash welfares for both governments). The politically optimal tariffs may fall outside the contract curve, though, if countries are sufficiently asymmetric.

2.2 Competitive Partial-Equilibrium Model of Trade Agreements

In the preceding discussion, we use a competitive general-equilibrium model and show that the purpose of the GATT/WTO is to provide a means of escape from a terms-of-trade driven Prisoners' Dilemma problem. As we discuss further in later sections, however, some design features of the GATT/WTO are most easily captured in the context of a simple partial-equilibrium model which includes specific parameters relating to political ingredients in governments' preferences. We thus now prepare for this analysis by presenting a simple partial-equilibrium model of trade. An additional and important benefit of our partial-equilibrium analysis is that it provides a concrete framework within which to understand and further explore the purpose of trade agreements. In particular, and as we show below, this analysis leads to new insights regarding the existence of international externalities that do not travel through the terms of trade when the set of trade-policy instruments is incomplete.

2.2.1 Partial-Equilibrium Model

We now present a two-country partial-equilibrium model of trade. The home country imports good x from the foreign country and exports good y to the foreign country. A numeraire good, n, is also traded between the two countries.[j]

Consumers in each country share a common utility function, which is quasilinear and additively separable across goods. The numeraire good enters the utility function in a linear way. Within any country, the demand for good i, where $i = x, y$, thus depends on the local price of good i relative to that of the numeraire good. Each good is supplied under conditions of perfect competition; hence, within a given country, the supply of good i, $i = x, y$, is also a function of the local price of good i relative to that of the numeraire good. As is standard, the numeraire good is produced in each country under constant returns to scale and is sufficiently abundant in each country that it is consumed in positive quantities. The numeraire good is also freely traded across countries so as to ensure that trade balance is achieved. We may normalize the price of the numeraire good to unity.[k]

[j] For further details regarding the specific model presented here, see Bagwell and Staiger (2001b).

[k] The model can be further specified with a description of the underlying factor market. Assume that the numeraire good is produced from labor alone, with one unit of labor generating one unit of the numeraire good. Suppose further that, in each country, the supply of labor is perfectly elastic at a unitary wage. A unitary wage is then implied. In each country, good i, $i = x, y$, is produced from labor alone subject to diminishing returns.

With the general-equilibrium trade-balance requirement achieved through trade in the numeraire good, whose price is fixed, a convenient feature of this model is that the market outcomes for good x are independent of those for good y. The model is thus partial equilibrium in nature.

We assume that consumers possess symmetric preferences across goods x and y. For $i = x, y$, each good i is then demanded in each country according to a symmetric demand function, D. Letting p_x and p_y denote the home-country local prices of goods x and y, respectively, we may represent the home-country demand function for good i as $D(p_i)$.[1] We may represent foreign demand functions similarly in terms of foreign local prices, p_x^* and p_y^*. To establish a basis for trade, we assume that the countries have different supply functions. The domestic supply functions are represented as $Q_x(p_x)$ and $Q_y(p_y)$, while the foreign supply functions are denoted as $Q_x^*(p_x^*)$ and $Q_y^*(p_y^*)$. For prices that are associated with strictly positive volumes, we assume that the demand function is strictly decreasing and the supply functions are increasing.[m] In line with our assumption that the home country imports good x, we assume that $Q_x(p) = Q_y^*(p) < Q_y(p) = Q_x^*(p)$ for any p such that $Q_y(p) > 0$. The associated competitive profit functions are represented by the strictly increasing functions, $\Pi_x(p_x)$, $\Pi_y(p_y)$, $\Pi_x^*(p_x^*)$, and $\Pi_y^*(p_y^*)$.

We assume that each country has a specific (ie, per unit) trade-policy instrument with which to affect the volume of trade for each good i, $i = x, y$. In particular, the home country has an import tariff, τ_x, that is applied to imports of good x and an export subsidy, τ_y, that is applied to exports of good y. An import subsidy is used if $\tau_x < 0$, while an export tax is applied when $\tau_y < 0$. The foreign country similarly has an import tariff, τ_y^*, and an export subsidy, τ_x^*. The "net tariff" for good x is then $\tau_x - \tau_x^*$, and similarly the net tariff for good y is captured as $\tau_y^* - \tau_y$.

We may now determine local prices as functions of tariffs. Provided that the net tariff is not so high as to prohibit trade, local prices in each country must obey the following arbitrage and market-clearing conditions:

$$p_i = p_i^* + \tau_i - \tau_i^* \tag{11}$$

$$Q_i(p_i) + Q_i^*(p_i^*) = D(p_i) + D(p_i^*) \text{ for } i = x, y. \tag{12}$$

For each good i, conditions (11) and (12) constitute two requirements on two prices; thus, we may solve these equations to determine the market-clearing local prices as functions of the associated net tariffs: $p_x(\tau_x - \tau_x^*)$, $p_x^*(\tau_x - \tau_x^*)$, $p_y^*(\tau_y^* - \tau_y)$ and $p_y(\tau_y^* - \tau_y)$. Under our assumptions, and provided that trade is not prohibited, it is direct to confirm

[1] Let the utility of the representative consumer take the form $u(C_x) + u(C_y) + C_n$, where C_i and C_n, respectively, represent consumption levels of good i, $i = x, y$, and the numeraire good n. The demand for good i is implicitly defined by the consumption level that satisfies the first-order condition: $u'(C_i) = p_i$.

[m] We allow that supply functions may be constant in order to include the possibility of endowment economies.

that an increase in the net tariff for a given good strictly increases (decreases) the local price in the importing (exporting) country. We thus have that $p_x'(\tau_x - \tau_x^*) > 0 > p_x^{*\prime}(\tau_x - \tau_x^*)$ and $p_y^{*\prime}(\tau_y^* - \tau_y) > 0 > p_y'(\tau_y^* - \tau_y)$.

We next consider the world (ie, offshore) prices. Let p_x^w and p_y^w denote the world prices for goods x and y, respectively. The world prices are defined as follows: $p_i \equiv p_i^w + \tau_i$ for $i = x, y$. For a given good, we may thus think of the world price as the price that prevails after the export policy is applied but before the import policy is imposed.[n] Given that we have already solved for the market-clearing local prices, we may use the definition of world prices to represent the market-clearing world prices as functions of tariffs: $p_x^w(\tau_x, \tau_x^*)$ and $p_y^w(\tau_y, \tau_y^*)$. Under our assumptions, if tariffs are not so high as to prohibit trade on a given good, then the world price is strictly decreasing in each argument: the world price strictly falls when the import tariff increases and when the export subsidy increases. In what follows, it is convenient to express market-clearing local prices in a country as explicit functions of the world prices and the country's own trade policy. For $i = x, y$, we thus define the market-clearing local price functions as $\hat{p}_i(\tau_i, p_i^w(\tau_i, \tau_i^*)) \equiv p_i^w(\tau_i, \tau_i^*) + \tau_i$ and $\hat{p}_i^*(\tau_i^*, p_i^w(\tau_i, \tau_i^*)) \equiv p_i^w(\tau_i, \tau_i^*) + \tau_i^*$.

It is instructive to consider the price implications of trade policies in more detail. Consider the home-country import good. An increase in the home-country import tariff causes the local price of the import good to rise in the home country; however, this price increase is less than one for one, since an increase in the import tariff reduces the world price and thereby the price that foreign exporters receive. When the home country raises its import tariff, it thereby shifts some of the cost of the tariff hike onto foreign exporters, who receive a lower price for their good and thus earn lower profits. For the partial-equilibrium model, the terms-of-trade externality that is associated with an import tariff increase thus admits a simple interpretation and corresponds to a reduction in the world price of the import good. Notice further than an increase in the foreign export subsidy causes the world price of the home import good to fall as well. Thus, an increase in an export subsidy generates a positive terms-of-trade externality for the importing country, by enabling this country to import at a lower world price.

We next identify the trade volumes that are implied by tariffs in this model. For the home country, we may define the import demand and export supply functions as $M(\hat{p}_x) \equiv D(\hat{p}_x) - Q_x(\hat{p}_x)$ and $E(\hat{p}_y) \equiv Q_y(\hat{p}_y) - D(\hat{p}_y)$, respectively. The foreign import demand and export supply functions are, respectively, defined in an exactly analogous manner as $M^*(\hat{p}_y^*) \equiv D(\hat{p}_y^*) - Q_y^*(\hat{p}_y^*)$ and $E^*(\hat{p}_x^*) \equiv Q_x^*(\hat{p}_x^*) - D(\hat{p}_x^*)$. Of course, under (12), the import demand for any good must equal the export supply for this good.

[n] Thus, for good x, $p_x \equiv p_x^w + \tau_x$. Similarly, for good y, we know from (11) that $p_y^* = p_y + \tau_y^* - \tau_y$. Using the definition of p_y^w, it follows that $p_y^* = p_y^w + \tau_y^*$.

Our assumptions also imply that, provided trade is not prohibited, the import demand (export supply) functions are strictly decreasing (increasing).

We now define government welfare functions. As Baldwin (1987) observes, a wide range of political-economy models are included under the assumption that governments maximize a weighted sum of consumer surplus, producer surplus and tariff revenue.[o] Let us use $\gamma_m \geq 1$ and $\gamma_e \geq 1$ to denote the weight that the domestic government attaches to the producer surplus earned by its import-competing and exporting firms, respectively. The welfare functions that the domestic government uses for its import and export goods, respectively, are then defined as follows:

$$W_x(\hat{p}_x, p_x^w) \equiv \int_{\hat{p}_x}^{\bar{p}} D(p_x)dp_x + \gamma_m \Pi_x(\hat{p}_x) + [\hat{p}_x - p_x^w]M(\hat{p}_x)$$

$$W_y(\hat{p}_y, p_y^w) \equiv \int_{\hat{p}_y}^{\bar{p}} D(p_y)dp_y + \gamma_e \Pi_y(\hat{p}_y) - [\hat{p}_y - p_y^w]E(\hat{p}_y),$$

where $\bar{p}$ denotes the reservation price for the demand function, D.[p] The home-government welfare function is then the sum of its welfare functions for import and export goods:

$$W(\hat{p}_x, \hat{p}_y, p_x^w, p_y^w) \equiv W_x(\hat{p}_x, p_x^w) + W_y(\hat{p}_y, p_y^w).$$

Welfare functions for the foreign government may be defined in an exactly analogous way. For given welfare weights γ_m^* and γ_e^*, the foreign-government import and export welfare functions are

$$W_y^*(\hat{p}_y^*, p_y^w) \equiv \int_{\hat{p}_y^*}^{\bar{p}} D(p_y^*)dp_y^* + \gamma_m^* \Pi_y^*(\hat{p}_y^*) + [\hat{p}_y^* - p_y^w]M^*(\hat{p}_y^*)$$

$$W_x^*(\hat{p}_x^*, p_x^w) \equiv \int_{\hat{p}_x^*}^{\bar{p}} D(p_x^*)dp_x^* + \gamma_e^* \Pi_x^*(\hat{p}_x^*) - [\hat{p}_x^* - p_x^w]E^*(\hat{p}_x^*).$$

The foreign-government welfare function is thus

$$W^*(\hat{p}_y^*, \hat{p}_x^*, p_y^w, p_x^w) \equiv W_y^*(\hat{p}_y^*, p_y^w) + W_x^*(\hat{p}_x^*, p_x^w).$$

For simplicity, in this section, we suppose further that the welfare weights are symmetric across governments, so that $\gamma_m = \gamma_m^* \geq 1$ and $\gamma_e = \gamma_e^* \geq 1$.

[o] The lobbying models of Grossman and Helpman (1994, 1995) provide microfoundations for such preferences.

[p] In terms of the utility function detailed in footnote l, we may understand that the tariff revenue term depicted in the welfare functions corresponds to consumer surplus that arises from the consumption of the numeraire good.

2.2.2 Prisoners' Dilemma

With the partial-equilibrium model now fully specified, we may briefly consider the basic Prisoners' Dilemma that arises in this model. We do so by considering the Nash, efficient and politically optimal tariffs in this model.

To characterize Nash trade policies, we first must consider the best-response or optimal trade policies for the domestic and foreign governments. The best-response policies are defined by

$$\begin{aligned}\text{Home}: W_{\hat{p}_x}\frac{d\hat{p}_x}{d\tau_x}+W_{p_x^w}\frac{\partial p_x^w}{\partial \tau_x}=0=W_{\hat{p}_y}\frac{d\hat{p}_y}{d\tau_y}+W_{p_y^w}\frac{\partial p_y^w}{\partial \tau_y}\\ \text{Foreign}: W^*_{\hat{p}^*_y}\frac{d\hat{p}^*_y}{d\tau^*_y}+W^*_{p_y^w}\frac{\partial p_y^w}{\partial \tau^*_y}=0=W^*_{\hat{p}^*_x}\frac{d\hat{p}^*_x}{d\tau^*_x}+W^*_{p_x^w}\frac{\partial p_x^w}{\partial \tau^*_x}.\end{aligned} \tag{13}$$

Thus, when setting its best-response trade policies, each government considers the effects of its policies on its country's local prices and terms of trade. A higher import tariff, for example, raises the local price of the import good. The local-price increase in turn implies a redistribution from consumer surplus to producer surplus on domestically produced units, which is especially attractive to the domestic government when γ_m is large. A higher local price for the import good, however, also leads to lower trade volume and an associated loss in tariff revenue. The terms-of-trade effect of a change in trade policy encourages the use of import tariffs and discourages the use of export subsidies. For example, a higher import tariff depresses the world price and thereby shifts foreign surplus into domestic tariff revenue.

The home-government best-response trade policies satisfy the first set of equations in (13) and may be represented as $\tau_x^R(\tau_x^*)$ and $\tau_y^R(\tau_y^*)$, and the best-response trade policies of the foreign government satisfy the second set of equations in (13) and are represented as $\tau_y^{*R}(\tau_y)$ and $\tau_x^{*R}(\tau_x)$. An *interior Nash equilibrium* vector of tariffs, $(\tau_x^N, \tau_y^N, \tau_y^{*N}, \tau_x^{*N})$, satisfies all of the requirements in (13), so that each trade policy of each government is a best response. As Dixit (1987) observes, an *autarky Nash equilibrium* also exists for tariff games. In the present model, as Bagwell and Staiger (2001b) confirm, if for a given good the import tariff is sufficiently high and the export tariff is also sufficiently high, then trade is prohibited and neither government has incentive to cut its tariff sufficiently so as to induce positive trade volume.

To analyze efficient tariffs, we first observe that governments have adequate instruments in this model to generate direct transfers. Let us suppose that the domestic government raises its import tariff and that the foreign government raises its export subsidy by the same amount. The net tariff on the home-country import good is then unaffected. This means that local prices and thus trade volume are unaffected by the proposed change in trade policy. The change does, however, generate a lower world price for the home-country import good. As is evident from the welfare functions above,

if $\hat{p}_x$ and $\hat{p}_x^*$ are unchanged and p_x^w falls by, say, one unit, then the domestic government enjoys a welfare gain in the amount of $M(\hat{p}_x)$ and the foreign government suffers a welfare loss in the amount $E^*(\hat{p}_x^*)$. Since the market-clearing condition (12) ensures that $M(\hat{p}_x) = E^*(\hat{p}_x^*)$, we may conclude that governments can make lump-sum government-to-government transfers through policy changes that preserve local prices and alter world prices.

This conclusion implies that a vector of tariffs is *efficient* if and only if it maximizes joint government welfare, $W(\hat{p}_x, \hat{p}_y, p_x^w, p_y^w) + W^*(\hat{p}_y^*, \hat{p}_x^*, p_y^w, p_x^w)$. Looking at the welfare functions defined above, we may also easily confirm that, at given local prices, joint government welfare is independent of world prices. Since joint welfare depends only on local prices, while local prices in turn depend only on net tariffs, we see that efficiency determines only the net tariff for each good. For example, if $\gamma_m = \gamma_x = 1$ so that each government maximizes national economic welfare, then efficiency requires that the net tariff is zero. For a given good, however, this requirement is met whenever the import tariff equals the export subsidy. If we start at free-trade policies and then increase the import tariff and export subsidy for this good in the same amount, then the outcome remains efficient. The joint increase in the import tariff and export subsidy, however, lowers the world price of the good and thus redistributes welfare from the exporting country to the importing country.

Given that joint government welfare depends only on net tariffs, we may fully characterize the set of efficient tariff vectors for any good by finding the net tariff for that good that maximizes joint government welfare. Furthermore, since the model is symmetric across goods, the efficient net tariff is common across goods. We may therefore focus on good x. Assuming that joint government welfare is strictly concave with respect to net tariffs, the associated first-order condition for an efficient tariff vector is

$$W_{\hat{p}_x} \frac{d\hat{p}_x}{d\tau_x} + W^*_{\hat{p}_x^*} \frac{d\hat{p}_x^*}{d\tau_x} = 0. \tag{14}$$

Thus, the set of efficient tariff vectors is the set of tariff vectors that deliver for each good the net tariff that satisfies (14). For any good, different combinations of tariffs that deliver the same net tariff vector simply indicate alternative international transfer patterns achieved through different world prices.

We now turn to the three questions raised at the start of the section. First, we consider the efficiency of Nash trade policies. We expect that Nash trade policies are inefficient, since governments are motivated by terms-of-trade considerations when setting Nash policies while efficiency imposes requirements only on local prices and thus the net tariff. We find that, at Nash policies,

$$W_{\hat{p}_x} \frac{d\hat{p}_x}{d\tau_x} + W^*_{\hat{p}_x^*} \frac{d\hat{p}_x^*}{d\tau_x} = W_{\hat{p}_x} \frac{d\hat{p}_x}{d\tau_x} - W^*_{\hat{p}_x^*} \frac{d\hat{p}_x^*}{d\tau_x^*} = -W_{p_x^w} \frac{\partial p_x^w}{\partial \tau_x} + W^*_{p_x^w} \frac{\partial p_x^w}{\partial \tau_x^*} < 0, \tag{15}$$

where the first equality follows since $\hat{p}_x^*$ ultimately depends on the net tariff, $\tau_x - \tau_x^*$, the second equality uses (13), and the inequality follows since $W^*_{p_x^w} = -W_{p_x^w} > 0$ and the world price is strictly decreasing in each of its arguments. Given our assumption that the joint government welfare function is strictly concave, (14) and (15) imply that the Nash net tariff is higher than efficient.

Second, we establish a general sense in which governments can achieve mutual gains relative to Nash policies only through reciprocal liberalization. In particular, suppose that governments start with Nash policies and then make mutual policy adjustments that change the net tariffs for goods x and y in a symmetric way. Importantly, while the net tariff changes under consideration are symmetric across goods, we are not assuming that the underlying policy adjustments are symmetric across governments. Since the net tariff under Nash policies is higher than efficient, total government welfare rises—and thus mutual gains to government welfare are feasible—only if the net tariff falls. Furthermore, given that each government selects its best-response policies at the Nash equilibrium, mutual gains are possible only if each government makes adjustments to its policies that contribute to some degree to the fall in net tariffs.

Finally, we confirm that terms-of-trade motivations provide the sole rationale for trade agreements in the partial-equilibrium model just as in the general-equilibrium model of the previous section. For the partial-equilibrium model, we follow Bagwell and Staiger (2001b) and define the *politically optimal* tariffs as the vector of tariffs satisfying

$$W_{\hat{p}_x} = W_{\hat{p}_y} = 0 = W^*_{\hat{p}_y^*} = W^*_{\hat{p}_x^*}. \tag{16}$$

The four requirements in (16) determine values for all four trade-policy instruments. In the symmetric model considered here, a common net tariff is thereby determined for both goods. Using (14) and (16), it is now clear that the net tariff is also efficient. In other words, the politically optimal tariffs are efficient. Following Bagwell and Staiger (2001b), we may interpret this finding to mean that the terms-of-trade externality is the sole rationale for a trade agreement in the partial-equilibrium model as well.

Bagwell and Staiger (2001b) also consider a "linear-quadratic" version of the above model in which demand is linear with a reservation price of unity and supply functions are linear. Specifically, they posit that $D(p) = \bar{p} - p$, $\bar{p} = 1$, $Q_y(p) = Q_x^*(p) = p$ and $Q_x(p) = Q_y^*(p) = p/2$. The associated profit functions are quadratic: $\Pi_y(p) = \Pi_x^*(p) = p^2/2$ and $\Pi_x(p) = \Pi_y^*(p) = p^2/4$.[q] If the political-economy weights are not

[q] Under the assumption that labor supply is perfectly elastic at a unitary wage, the domestic supply and profit functions can be derived from the following production functions: $Q_x = \sqrt{L_x}$ and $Q_y = \sqrt{2L_y}$, where L_i units of labor are used in the production of good i, $i = x, y$. Analogous remarks apply to the foreign functions. The demand functions can be derived from a representative consumer with utility function $U = C_n + (C_x - C_x^2/2) + (C_y - C_y^2/2)$, where C_n and C_i denote consumption of the numeraire good and good i, $i = x, y$.

too large, so that $\gamma_m \in [1, 3]$ and $\gamma_e \in [1, 3]$, then second-order conditions hold and best-response trade-policy functions are strictly increasing. Intuitively, when the foreign export subsidy increases, the home country imports a larger volume and domestic production falls. The former effect magnifies the terms-of-trade benefit of raising the import tariff. The latter effect, however, weakens the political-economy benefit of using a higher import tariff to raise the local price, since the higher price would apply to a smaller volume of domestic production. The finding that trade policies are strategic complements indicates that, in the linear model at least, the magnification of the terms-of-trade effect dominates the weakening of the political-economy effect.

Under the maintained assumption that the political weight on the import-competing sector is not too large relative to that on the export sector, in the specific sense that $5/8 + 9\gamma_e/8 > \gamma_m$, Bagwell and Staiger (2001b) establish that the interior Nash equilibrium is unique and has positive trade volume. The Nash import tariff is always positive; however, the sign of the export subsidy is ambiguous.[r] Intuitively, the Nash import tariff is strictly positive even for a government that maximizes national economic welfare, since the government of a large country can use a positive import tariff to improve its country's terms of trade. If $\gamma_m > 1$ so that the government receives political benefits from raising the local price of the import good, then the case for a positive Nash import tariff is reinforced. An export tax is likewise optimal for a government that maximizes national economic welfare; however, the Nash export subsidy is positive when γ_e is sufficiently large, since the government then enjoys a large benefit from using an export subsidy to elevate the local price of the export good. Bagwell and Staiger also report that the politically optimal import tariff and export subsidy are nonnegative and correspond to free trade when political economy effects are absent.[s] Finally, they confirm that joint government welfare is strictly concave in the net tariff and that the Nash trade policies are too restrictive, as the trade volume in the interior Nash equilibrium is strictly less than would be efficient for the governments.

The partial-equilibrium model presented here provides a simple and concrete framework with which to understand and explore the terms-of-trade driven Prisoners' Dilemma problem. Like the general-equilibrium model presented above, the partial-equilibrium model indicates that a mutually beneficial trade agreement between governments must entail reciprocal reductions in trade barriers. Thus, the partial- and general-equilibrium frameworks generate the same broad conclusions regarding the purpose of trade agreements.

[r] Grossman and Helpman (1995) present a related finding in a more sophisticated model in which political-economic preferences are endogenously determined by lobbying of politically organized sectors. In their model, the ad valorem import tariff is positive for a good with an organized import-competing industry, and the ad valorem export subsidy is negative (ie, an export tax is used) for a good in an export sector that is not organized. In other cases, the sign of the optimal policy is ambiguous, since terms-of-trade and political considerations pull in opposite directions.

[s] The politically optimal import tariff and export subsidy are zero when $\gamma_m = 1$ and $\gamma_e = 1$, respectively.

2.2.3 Missing Instruments in the Partial-Equilibrium Model

We now consider the purpose of trade agreements when the partial-equilibrium model is modified so that some trade-policy instruments are "missing." Recall that the politically optimal trade-policy vector is a vector with four policies—an import tariff and an export subsidy for each country—such that each country achieves its preferred local price on each good: $W_{\hat{p}_x} = W_{\hat{p}_y} = 0 = W^*_{\hat{p}^*_y} = W^*_{\hat{p}^*_x}$. Intuitively, if policies satisfy these requirements, then a small change in any one trade policy cannot generate a Pareto improvement, since the first-order welfare effects of the resulting local-price changes are zero and since the resulting change in any world price constitutes a pure transfer from one government to the other and thus cannot be a source of mutual gain.

Suppose now, however, that export policies are prohibited. In this restricted-instrument setting, let us assume that each government selects its import tariff in a politically optimal fashion. The resulting politically optimal import tariffs are those tariffs that deliver $W_{\hat{p}_x} = 0$ and $W^*_{\hat{p}^*_y} = 0$ when all export subsidies are set to zero. In general, the politically optimal import tariffs for the restricted-instrument setting differ from the politically optimal import tariffs for the setting in which both import and export instruments are available.[t] When the politically optimal import tariffs for the restricted-instrument setting are used, we may easily verify that $W_{\hat{p}_y} > 0$ and $W^*_{\hat{p}^*_x} > 0$ when $\gamma_e > 1$. Thus, when these import tariffs are used and $\gamma_e > 1$, each government would prefer a higher local price for its export good. A government is unable to deliver such a price on its own, however, when export subsidies are prohibited.

We now claim that the resulting four-tuple of policies, where import tariffs are politically optimal in the restricted-instrument setting and export subsidies are zero, is inefficient. To establish this claim, we need only note that each government would strictly gain if the governments were to reciprocally and symmetrically exchange small import tariff cuts. A small import tariff cut by the home government would change the local price of good x in the home country, but this has no first-order effect on the welfare of the home government since $W_{\hat{p}_x} = 0$. The tariff cut would also raise the world price and thus the local price of the good in the foreign country; furthermore, when $\gamma_e > 1$, this local-price effect would generate a first-order welfare gain for the foreign government since $W^*_{\hat{p}^*_x} > 0$. Symmetrically, a small import tariff cut by the foreign government would change the local prices of good y in a manner that would generate no first-order effect on the welfare of the foreign government (since $W^*_{\hat{p}^*_y} = 0$) and a first-order gain in the welfare of the domestic government when $\gamma_e > 1$ (since $W_{\hat{p}_y} > 0$). Finally, each tariff

[t] When the foreign government provides an export subsidy at level τ^*_x, the import tariff for the home government that satisfies $W_{\hat{p}_x} = 0$ is given by $\tau_x = 4(\gamma_m - 1)(1 - \tau^*_x)/(25 - 4\gamma_m)$. Thus, if $\gamma_m > 1$, then the politically optimal selection for the home import tariff differs depending on whether export subsidies are prohibited ($\tau^*_x = 0$) or the foreign export subsidy is set at its politically optimal level ($\tau^*_x > 0$ when $\gamma_e > 1$).

cut would induce a terms-of-trade loss for the importing country and a terms-of-trade gain for the exporting country. In a symmetric setting, a country imports as much as it exports, and the overall welfare effect for each government of the resulting changes in world prices is zero.[u] It follows that each government gains from a small and symmetric exchange of reciprocal tariff cuts.

In the restricted-instrument setting, therefore, when governments are not motivated by the terms-of-trade implications of their trade-policy selections, a problem may remain for a trade agreement to solve. The problem is fundamentally a missing-instrument problem. When $\gamma_e > 1$, a domestic government that ignores the terms-of-trade implications of its policies would like to subsidize its exports so as to achieve its preferred local price. Since the domestic government does not have an instrument with which to perform this subsidization, the local price for its export good may be lower than it prefers. A *local-price externality* then arises: an import tariff reduction by the foreign government can generate a positive international externality for the domestic government by raising the local price of the domestic export good. Interestingly, mutual improvements for governments are again achieved in this setting through reciprocal reductions in import tariffs. It is also interesting to note that the local-price externality arises in the competitive setting, even though the domestic government then has no direct interest in the foreign local prices. As we argue below, related findings occur in imperfect competition settings, although in that case the domestic government also has a direct interest in foreign local prices.

2.3 Monopolistic Competition Model of Trade Agreements

Our discussion to this point diagnoses the purpose of a trade agreement from the perspective of models that assume perfectly competitive markets. A large literature, however, suggests that unilateral trade policies have novel implications in imperfectly competitive markets.[v] We thus turn our attention next to the possible implications of imperfectly competitive markets for the purpose of trade agreements. To focus our discussion, we develop a variant of the monopolistic competition model used by Venables (1987). The variant that we explore is also examined by Helpman and Krugman (1989) and further developed by Bagwell and Staiger (2015).[w] The model features integrated markets, "iceberg" transport costs, monopolistically competitive firms which produce differentiated varieties under increasing returns, CES preferences for differentiated products, free entry in both the home and foreign countries, and a homogeneous "outside" good that is produced under constant returns to scale. As before, we first develop the model and then

[u] When export subsidies are set equal to zero, the local price of the export good is the same as the world price of this good. It is nevertheless useful to break this price change into local- and world-price components, as captured for general trade policies in the welfare functions above.

[v] See, eg, Brander (1995) and Helpman and Krugman (1989).

[w] Our presentation here is closely related to that found in Bagwell and Staiger (2015).

consider the three questions raised at the start of the section. We conclude with a brief discussion of related research that uses other models of imperfect competition.

2.3.1 Monopolistic Competition Model

We again consider a two-country model. The utility functions for consumers in the home and foreign countries are symmetric and are, respectively, given by

$$\begin{aligned} U &= (C_D)^{\theta}\left(\frac{1}{\theta}\right) + C_Y \\ U^* &= (C_D^*)^{\theta}\left(\frac{1}{\theta}\right) + C_Y^*, \end{aligned} \tag{17}$$

where $\theta \in (0, 1)$, C_D is a home-country index of consumption of a bundle of differentiated goods, C_Y is home-country consumption of a homogeneous good Y, and where C_D^* and C_Y^* are defined analogously in the foreign country. At this point, we respectively treat C_D and C_D^* as home and foreign consumption of a single composite good, which we refer to as good D. We assume that good Y is a numeraire good that is produced in each country from labor, where each unit of labor produces a single unit of good Y. The home and foreign countries, respectively, are endowed with large supplies of labor, L and L^*. Each country then produces good Y, which is freely traded across countries. Under these assumptions, we may fix the wage and thus the price of good Y in each country at unity.

In each country, utility maximization establishes an equality between the ratio of marginal utilities across goods D and Y and the ratio of prices between goods D and Y, where we recall that the price of good Y is unity. Letting P and P^*, respectively, denote the price of good D in the home and foreign countries, we thus have that

$$\begin{aligned} C_D &= P^{-\epsilon} \\ C_D^* &= (P^*)^{-\epsilon} \end{aligned} \tag{18}$$

where in each country $\epsilon \equiv 1/(1-\theta) > 1$ is the price elasticity of demand for good D. Letting I and I^* denote income in the home and foreign country, respectively, we may use (17) and (18) to derive the home- and foreign-country indirect utility functions:

$$\begin{aligned} V(P, I) &= P^{-\epsilon\theta}\left(\frac{1}{\epsilon\theta}\right) + I \\ V^*(P^*, I^*) &= (P^*)^{-\epsilon\theta}\left(\frac{1}{\epsilon\theta}\right) + I^* \end{aligned} \tag{19}$$

where income is measured in terms of the numeraire good Y.

We assume now that C_D is a consumption index that takes a CES form. Specifically, we assume that $C_D = [\sum_i (c_i)^{\alpha}]^{1/\alpha}$, where $\alpha \in (0, 1)$ and c_i is the consumption level in the home country of variety i of the differentiated good. We may similarly represent

$C_D^* = [\sum_i (c_i^*)^\alpha]^{1/\alpha}$, where c_i^* is the consumption level in the foreign country of variety i of the differentiated good. Letting p^i and p^{*i} denote the prices of variety i in the home and foreign countries, respectively, it can be shown (see Dixit and Stiglitz, 1977) that

$$\begin{aligned} P &= \left[\sum_i (p^i)^{\alpha/(\alpha-1)}\right]^{(\alpha-1)/\alpha} \\ P^* &= \left[\sum_i (p^{*i})^{\alpha/(\alpha-1)}\right]^{(\alpha-1)/\alpha} \end{aligned} \tag{20}$$

are the associated home- and foreign-country price indices for good D.

Letting $\sigma \equiv 1/(1-\alpha) > 1$, we may now represent the demand for variety i in the home and foreign countries, respectively, as

$$\begin{aligned} c^i &= C_D \left(\frac{p^i}{P}\right)^{-\sigma} \\ c^{*i} &= C_D^* \left(\frac{p^{*i}}{P^*}\right)^{-\sigma}. \end{aligned} \tag{21}$$

Combining (18) with (21) yields

$$\begin{aligned} c^i &= (p^i)^{-\sigma} P^{\sigma-\epsilon} \equiv c^i(p^i, P) \\ c^{*i} &= (p^{*i})^{-\sigma} (P^*)^{\sigma-\epsilon} \equiv c^{*i}(p^{*i}, P^*), \end{aligned} \tag{22}$$

which completes the description of the demand side.

We now consider the costs confronted by firms. Production of any variety i entails a fixed cost of labor, $F > 0$, and a constant marginal cost of labor, $\lambda > 0$. The positive fixed cost ensures that no variety is produced by more than one firm; thus, if variety i is produced somewhere, then exactly one firm in the world supplies this variety. In addition, firms confront trade costs. Let $\phi > 0$ denote the "iceberg" transport cost associated with international trade. An exporting firm also faces trade costs in the form of ad valorem export and import tariffs. Let the home-country ad valorem import and export tariffs, respectively, be denoted as τ_h and τ_h^*, and let the foreign-country ad valorem import and export tariffs similarly be denoted as τ_f^* and τ_f, respectively. Notice that an export subsidy by the home country, for example, is captured by a negative value for τ_h^*.

We assume that the markets are integrated and that tariffs are not prohibitive. It then follows that price wedges across countries for any given variety are determined by the total trade costs. Letting $\iota \equiv 1 + \phi + \tau_h + \tau_f$ and $\iota^* \equiv 1 + \phi + \tau_f^* + \tau_h^*$, we have that

$$\begin{aligned} p_h^{*i} &= \iota^* p_h^i \\ p_f^i &= \iota p_f^{*i}, \end{aligned} \tag{23}$$

where p_h^i and p_h^{*i} are the respective prices of home-produced variety i in the home and foreign countries and where p_f^{*i} and p_f^i are the respective prices of foreign-produced variety i in the foreign and home countries.[x]

The profit for a home-country firm that produces variety i is now given by

$$\pi^i = (p_h^i - \lambda)[c^i(p_h^i, P) + (1+\phi)c^{*i}(p_h^{*i}, P^*)] - F, \tag{24}$$

where the number of varieties in each country enters profit only through the price indices.[y] Each firm selects its profit-maximizing price while taking as given the price indices, P and P^*. Using (22), (23) and (24), the profit-maximizing prices for home-produced variety i are

$$\begin{aligned} p_h^i &= \frac{\sigma}{\sigma - 1}\lambda \equiv \hat{p} \\ p_h^{*i} &= \iota^* \hat{p} \equiv p_h^*(\iota^*). \end{aligned} \tag{25}$$

Similarly, the profit for a foreign-country firm that produces variety i is given by

$$\pi^{*i} = (p_f^{*i} - \lambda)[c^{*i}(p_f^{*i}, P^*) + (1+\phi)c^i(p_f^i, P)] - F, \tag{26}$$

and the profit-maximizing prices are

$$\begin{aligned} p_f^{*i} &= \frac{\sigma}{\sigma - 1}\lambda \equiv \hat{p} \\ p_f^i &= \iota \hat{p} \equiv p_f(\iota). \end{aligned} \tag{27}$$

Our next step is to represent the price indices that result from profit-maximizing pricing. Using (20), (25) and (27), we obtain

$$\begin{aligned} P &= [n_h(\hat{p})^{\alpha/(\alpha-1)} + n_f(p_f)^{\alpha/(\alpha-1)}]^{(\alpha-1)/\alpha} \equiv P(n_h, n_f, p_f) \\ P^* &= [n_f(\hat{p})^{\alpha/(\alpha-1)} + n_h(p_h^*)^{\alpha/(\alpha-1)}]^{(\alpha-1)/\alpha} \equiv P^*(n_h, n_f, p_h^*), \end{aligned} \tag{28}$$

where n_h and n_f are the number of firms producing differentiated varieties in the home and foreign countries, respectively, and where we suppress the dependence of p_f on ι and of p_h^* on ι^* here and henceforth. Using (28), we observe that, for given n_h and n_f, the price index in any one country is rising in the total trade costs faced by exporters from the other country but is independent of the total trade costs faced by its own

[x] This formulation assumes that a foreign importer of variety i, for example, purchases at the home-country factory gate at price p_h^i and then pays the export tax, $\tau_h^* p_h^i$, import tax, $\tau_f^* p_h^i$, and transport cost, ϕp_h^i.

[y] This expression reflects the fact that a foreign importer must purchase $1 + \phi$ units at the home-country factory gate in order that 1 unit may be delivered for consumption in the foreign country. Our formulation assumes that international shipping services are freely traded.

exporters. It is also interesting to consider the implications of a reallocation of firms across countries. Referring again to (28), if the total trade costs are positive (ie, $\iota > 1$ and $\iota^* > 1$), then an increase in n_h that is matched by a reduction in n_f results in a fall in the home-country price index P and a rise in the foreign-country price index P^*.

We are now prepared to determine n_h and n_f using free-entry conditions. The free-entry conditions are defined as follows:

$$\begin{aligned} c(\hat{p}, P(n_h, n_f, p_f)) + (1+\phi)c^*(p_h^*, P^*(n_h, n_f, p_h^*)) &= F/(\hat{p}-\lambda) \\ c^*(\hat{p}, P^*(n_h, n_f, p_h^*)) + (1+\phi)c(p_f, P(n_h, n_f, p_f)) &= F/(\hat{p}-\lambda), \end{aligned} \tag{29}$$

where we exploit the symmetry of the differentiated sector and now eliminate variety i superscripts. The resulting values for n_h and n_f may be expressed as $n_h(p_f, p_h^*)$ and $n_f(p_f, p_h^*)$. After inserting these expressions into (28), we may abuse notation somewhat and write the free-entry values for the price indices as $P(p_f, p_h^*)$ and $P^*(p_f, p_h^*)$.

As first noted by Venables (1987), the monopolistic competition model exhibits an interesting "firm-delocation effect" from trade protection. To illustrate the point, suppose that we begin at global free trade and then introduce a small and positive import tariff in the home country (ie, $\tau_h > 0$). We see from (25) and (27) that p_h^* is unaltered while p_f rises; thus, for given values of n_h and n_f, it follows from (28) that P^* is unchanged and that P rises. This in turn implies that $c(\hat{p}, P(n_h, n_f, p_f))$ and thus the LHS of the first expression in (29) rises. In addition, since it can be confirmed that p_f rises more than does P, it follows as well that $c(p_f, P(n_h, n_f, p_f))$ and thus the LHS of the second expression in (29) falls. Therefore, for given values of n_h and n_f, the introduction of a small and positive import tariff in the home country generates positive profits for home firms and negative profits for foreign firms. The restoration of zero profits in each country then requires a decrease in P and an increase in P^*.[z] The final step is to note from (28) that these changes in the price indices in turn require an increase in n_h and a decrease in n_f, that is, a "delocation" of firms from the foreign to the home country. A related argument establishes that the introduction of a small export subsidy in the foreign country (ie, $\tau_f < 0$) leads to a rise in P and a fall in P^*, which in turn require a decrease in n_h and an increase in n_f, that is, a "delocation" of firms from the home to the foreign country.

More generally, the firm-delocation effect indicates that a slight increase in the home-country import tariff "delocates" foreign firms to the home-country market, which benefits home-country consumers who enjoy the lower price index that derives

[z] Intuitively, home and foreign profits can adjust in different directions only if the price indices also move in different directions; furthermore, given the positive trade cost $\phi > 0$ and the symmetry of the model, local sales contribute relatively more to the profits of each country's firms than do export sales. See Bagwell and Staiger (2015) and Helpman and Krugman (1989) for additional discussion.

from reduced trade costs and harms foreign-country consumers who experience increased trade costs and thus a higher price index. By a similar logic, a slight increase in the foreign-country export subsidy "delocates" home firms to the foreign-country market, which benefits foreign-country consumers while harming home-country consumers.

As our preceding discussion suggests, trade policy affects welfare through the firm-delocation effect and the associated impact on price indices. But trade policy may also generate tariff revenue and impact income and thus welfare through this channel as well. In the model that we consider here, where profits are zero due to free entry, income in a given country is given by its labor force (which receives a wage of unity) plus the country's net trade tax revenue. To formally represent income, we use M to denote the imports in the home country and E to represent exports from the home country, where by market-clearing home-country imports (exports) are equal to foreign-country exports (imports). We have that

$$\begin{aligned} M &= n_f(p_f, p_h^*)c(p_f, P(p_f, p_h^*)) \equiv M(p_f, p_h^*) \\ E &= n_h(p_f, p_h^*)c^*(p_h^*, P^*(p_f, p_h^*)) \equiv E(p_f, p_h^*), \end{aligned} \tag{30}$$

and so income levels in the home and foreign countries are given as[aa]

$$\begin{aligned} I &= L + \tau_h^* \hat{p} E(p_f, p_h^*) + \tau_h \hat{p} M(p_f, p_h^*) \\ I^* &= L^* + \tau_f \hat{p} M(p_f, p_h^*) + \tau_f^* \hat{p} E(p_f, p_h^*). \end{aligned} \tag{31}$$

To express welfare as a function of local and world prices, we must first define world prices. Consider first varieties that are exported from the home to the foreign country. We define the world price for these varieties as

$$p^{*w} = (1 + \tau_h^*)\hat{p} \equiv p^{*w}(\tau_h^*), \tag{32}$$

from which it follows that $\tau_h^* \hat{p} = p^{*w} - \hat{p}$. Using also that $p_h^* = \iota^* \hat{p}$ from (25), we then further have that $\tau_f^* \hat{p} = p_h^* - \phi \hat{p} - p^{*w}$. Consider now varieties that are exported from the foreign to the home country. The world price for these varieties is defined as

$$p^w = (1 + \tau_f)\hat{p} \equiv p^w(\tau_f), \tag{33}$$

from which we see that $\tau_f \hat{p} = p^w - \hat{p}$. We may use $p_f = \iota \hat{p}$ from (27) to further derive that $\tau_h \hat{p} = p_f - \phi \hat{p} - p^w$.

With these pricing relationships in hand, we may rewrite the expressions for income in (31) as

[aa] As indicated in footnote y, we assume that international shipping services are freely traded and earn zero profit.

$$
\begin{aligned}
I &= L + [p^{*w} - \hat{p}]E(p_f, p_h^*) + [p_f - \phi\hat{p} - p^w]M(p_f, p_h^*) \equiv I(p_h^*, p_f, p^w, p^{*w}) \\
I^* &= L^* + [p^w - \hat{p}]M(p_f, p_h^*) + [p_h^* - \phi\hat{p} - p^{*w}]E(p_f, p_h^*) \equiv I^*(p_h^*, p_f, p^w, p^{*w}).
\end{aligned} \tag{34}
$$

It is instructive to compare the expressions for incomes given by (31) and (34). The advantage of the expressions for incomes in (34) is that the price channels through which trade policies transmit international externalities are directly identified.

With (34) in place, we may now refer to (19) and (with some abuse in notation) write each country's indirect utility function in terms of local and world prices:

$$
\begin{aligned}
V(p_h^*, p_f, p^w, p^{*w}) &= P(p_f, p_h^*)^{-\epsilon\theta}\left(\frac{1}{\epsilon\theta}\right) + I(p_h^*, p_f, p^w, p^{*w}) \\
V^*(p_h^*, p_f, p^w, p^{*w}) &= P^*(p_f, p_h^*)^{-\epsilon\theta}\left(\frac{1}{\epsilon\theta}\right) + I^*(p_h^*, p_f, p^w, p^{*w}),
\end{aligned} \tag{35}
$$

where as noted earlier we may express price indices as functions of p_f and p_h^*. Since free entry ensures that profits are zero in this model, an obvious role for political motivations does not arise. We thus associate a country's welfare with its indirect utility function.

The monopolistic competition model considered here admits an interesting pattern of international externalities. As Helpman and Krugman (1989) note, one special feature of the model is that a country's import tariff does not generate an international externality that travels through world prices. For example, we may use (32) and (33) to confirm that the import tariff of the home country, τ_h, affects neither p^{*w} nor, more surprisingly, p^w. As Helpman and Krugman (1989) discuss, this feature arises because of CES preferences (which result in a constant markup) and the existence of a freely traded outside good (which ensures that the marginal cost λ is constant). By contrast, and as Bagwell and Staiger (2015) emphasize, these same features imply that a country's export tariff has an extreme impact on world prices, in that 100% of an export tariff is passed through to consumers abroad. For instance, using (32) and (33), we see that the export tariff of the home country, τ_h^*, has no effect on p^w but is fully passed through in p^{*w}.

We note, too, that the pattern of international externalities is more complicated in the monopolistic competition model than in the perfectly competitive models considered above. In the models with perfect competition, the welfare function of the home government, for example, is determined by the local and world prices, where the relevant local prices are those that prevail in the home country. In the model with monopolistic competition, however, and as (35) confirms, the home-country welfare depends directly also on the price of home-produced varieties in the foreign market (ie, on $p_h^* = p_h^*(\iota^*)$). Thus, the monopolistic-competition model admits a richer set of international externality channels. The key question for our purposes here, however, is whether these new channels lead to new rationales for a trade agreement. This question is addressed below.

2.3.2 Prisoners' Dilemma

We now briefly consider the basic Prisoners' Dilemma that arises in the monopolistic competition model. As with the models above, we characterize the Nash, efficient and politically optimal tariffs.

To characterize Nash trade policies, we begin by representing the best-response or optimal trade policies for the home and foreign governments. We assume in this discussion that the relevant second-order conditions hold. The home-country best-response import and export policies are then determined by the following two first-order conditions:

$$\begin{aligned} V_{p_f}\frac{dp_f}{d\iota} &= 0 \\ V_{p_h^*}\frac{dp_h^*}{d\iota^*} + V_{p^{*w}}\frac{dp^{*w}}{d\tau_h^*} &= 0, \end{aligned} \tag{36}$$

where the home-country import tariff τ_h affects p_f via ι, the home-country export tariff affects p_h^* via ι^*, and the indirect utility function V is given in (35). Similarly, the foreign-country best-response import and export policies satisfy the first-order conditions

$$\begin{aligned} V_{p_h^*}^*\frac{dp_h^*}{d\iota^*} &= 0 \\ V_{p_f}^*\frac{dp_f}{d\iota} + V_{p^w}^*\frac{dp^w}{d\tau_f} &= 0. \end{aligned} \tag{37}$$

We may now define the *Nash trade policies*, $(\tau_h^N, \tau_h^{*N}, \tau_f^{*N}, \tau_f^N)$, as the tariffs that simultaneously satisfy (36) and (37).

Consistent with our discussion earlier, we note that terms-of-trade effects are absent from the conditions that determine the best-response import policies (the top expressions in (36) and (37)) but present in the conditions that determine the best-response export policies (the bottom expressions in (36) and (37)). As Bagwell and Staiger (2015) discuss further, the optimal export policy for a given country thus represents a balance between the terms-of-trade gain that is associated with an export tariff and the firm-delocation benefit that is associated with an export subsidy.

Our next goal is to characterize efficient trade policies. Since this model also has sufficient trade-policy instruments with which to effect lump-sum transfers, efficient trade policies are those which maximize $V + V^*$. Using (34), we observe that total income can be written as

$$I(\cdot) + I^*(\cdot) = L + L^* + [p_f - \phi\hat{p} - \hat{p}]M(p_f, p_h^*) + [p_h^* - \phi\hat{p} - \hat{p}]E(p_f, p_h^*) \equiv T(p_f, p_h^*). \tag{38}$$

Notice in particular that world prices do not affect total income, which is to say that we can express total income as the function $T(p_f, p_h^*)$. Using (35) and (38), we can now also represent joint welfare as a function $J(p_f, p_h^*)$ where

$$V(\cdot) + V^*(\cdot) = P(p_f, p_h^*)^{-\epsilon\theta}\left(\frac{1}{\epsilon\theta}\right) + P^*(p_f, p_h^*)^{-\epsilon\theta}\left(\frac{1}{\epsilon\theta}\right) + T(p_f, p_h^*) \equiv J(p_f, p_h^*). \quad (39)$$

As this expression confirms, any trade-policy induced change in world prices corresponds simply to pure international rent shifting and does not affect efficiency.

Efficient trade policies thus maximize the joint welfare function $J(p_f, p_h^*)$ given in (39). By (25) and (27), respectively, we also know that $p_h^* = p_h^*(\iota^*)$ and $p_f = p_f(\iota)$; thus, joint welfare is a function of the four tariffs only through ι and ι^*. Assuming that the joint welfare function is strictly concave when treated as a function of ι and ι^*, the set of *efficient tariffs* is thus characterized by the following two first-order conditions:

$$\begin{aligned}[V_{p_h^*} + V_{p_h^*}^*]\frac{dp_h^*}{d\iota^*} &= 0\\ [V_{p_f} + V_{p_f}^*]\frac{dp_f}{d\iota} &= 0.\end{aligned} \quad (40)$$

Bagwell and Staiger (2015) further explore the two conditions in (40) and show that efficiency requires a net subsidy to trade along each trade channel (ie, $\tau_f^* + \tau_h^* < 0$ and $\tau_h + \tau_f < 0$).[ab] The intuition is that a net subsidy is desirable due to the positive markup in the differentiated sector.

We now turn again to the three questions raised at the start of the section. First, we consider the efficiency of Nash trade policies. Once again, we expect that Nash trade policies are inefficient, since governments are motivated by world-price considerations when setting Nash policies but not when maximizing joint welfare. More formally, after adding the bottom Nash condition in (36) to the top Nash condition in (37), and likewise adding the top Nash condition in (36) to the bottom Nash condition in (37), we arrive at the following:

$$\begin{aligned}[V_{p_h^*} + V_{p_h^*}^*]\frac{dp_h^*}{d\iota^*} &= -E\frac{dp^{*w}}{d\tau_h^*} < 0\\ [V_{p_f} + V_{p_f}^*]\frac{dp_f}{d\iota} &= -M\frac{dp^w}{d\tau_f} < 0\end{aligned} \quad (41)$$

[ab] See Helpman and Krugman (1989) for related discussion.

where we use (34) and (35) to obtain that $V_{p^{*w}} = E$ and $V_{p^w}^* = M$. Comparing (40) and (41), it is now immediate that Nash trade policies are inefficient.

Given the assumed second-order conditions and the symmetry of the model, we know that Nash and efficient tariffs are each such that the total tariffs satisfy $\iota = \iota^*$. Starting at the Nash equilibrium, if we were to undertake any change in underlying tariffs that delivered a symmetric increase in $\iota = \iota^*$, then the change in joint welfare would be given by the sum of the terms on the LHS of the equalities in (41), when evaluated at $\iota^N \equiv 1 + \phi + \tau_h^N + \tau_f^N = 1 + \phi + \tau_f^{*N} + \tau_h^{*N} \equiv \iota^{*N}$. As is evident from (41), starting at the Nash equilibrium, a symmetric change in $\iota = \iota^*$ increases joint welfare if and only if $\iota = \iota^*$ is decreased. It thus follows that the total trade cost, $\iota = \iota^*$, is strictly higher at Nash tariffs than at efficient tariffs.

We turn now to our second question and explore whether a mutually beneficial trade agreement requires reciprocal trade liberalization. Our preceding discussion already establishes that a trade agreement that delivers symmetric changes in total trade costs can generate greater joint welfare if and only if total tariffs are reduced from Nash levels. Thus, at least in the context of trade policy adjustments that maintain symmetric total trade costs, $\iota = \iota^*$, mutual gains are possible starting at Nash only if reciprocal trade liberalization occurs in the sense that $\tau_h + \tau_f = \tau_f^* + \tau_h^*$ is reduced. Just as in our discussion of the partial-equilibrium model with perfect competition, such efficiency-enhancing paths may involve adjustments in underlying tariffs that are asymmetric across countries. Mutual gains are again possible, however, only if each government makes adjustments to its policies that contribute to some degree to the fall in total tariffs.

Finally, we consider whether terms-of-trade motivations represent the sole rationale for trade agreements in the monopolistic competition model. To explore this issue, we again define the politically optimal tariffs to be those tariffs that hypothetically would be chosen by governments unilaterally if they did not value the pure international rent-shifting associated with the terms-of-trade movements induced by their unilateral tariff choices. For the monopolistic competition model under consideration here, when making their respective politically optimal tariff selections, the home-country government acts as if $V_{p^{*w}} = 0 = V_{p^w}$ while the foreign-country governments acts as if $V_{p^w}^* = 0 = V_{p^{*w}}^*$.[ac] Formally, and following Bagwell and Staiger (2015), we define the *politically optimal tariffs* for the monopolistic competition model as the vector of tariffs satisfying

$$V_{p_f} \frac{dp_f}{d\iota} = V_{p_h^*} \frac{dp_h^*}{d\iota^*} = 0 = V_{p_h^*}^* \frac{dp_h^*}{d\iota^*} = V_{p_f}^* \frac{dp_f}{d\iota}. \tag{42}$$

[ac] As noted previously, in the monopolistic competition model considered here, a country cannot use its import policy to change the world price of varieties produced abroad. For consistency, in our definition of politically optimal tariffs, we nevertheless include the requirement that such a change would not be valued.

The four equations in (42) determine the politically optimal tariff vector, $(\tau_h^{PO}, \tau_h^{*PO}, \tau_f^{*PO}, \tau_f^{PO})$. In the symmetric model considered here, a common total tariff is determined for each direction of trade, $\tau_h^{PO} + \tau_f^{PO} = \tau_f^{*PO} + \tau_h^{*PO}$, so that the resulting total trade cost is also symmetric: $\iota^{PO} \equiv 1 + \phi + \tau_h^{PO} + \tau_f^{PO} = 1 + \phi + \tau_f^{*PO} + \tau_h^{*PO} \equiv \iota^{*PO}$. Using (40) and (42), it is now immediate that the politically optimal tariffs are efficient. Thus, in the monopolistic competition model as well, the terms-of-trade externality is the sole rationale for a trade agreement.

Intuitively, in the Nash equilibrium of the monopolistic competition model, each government is mindful of the beneficial firm-delocation effect that import tariffs and export subsidies offer, and each government is also attentive to the terms-of-trade gain that export tariffs provide. The effects of trade policies on trade volumes and thereby tariff revenue are also considered. By contrast, when governments select politically optimal policies, they ignore the terms-of-trade impacts of trade (namely, export) policies and focus on the local-price implications of trade policies. The local prices that can be influenced by trade policy in this model are the domestic prices of varieties produced abroad, p_h^* and p_f, where these prices in turn are determined by total trade costs, $\iota = 1 + \phi + \tau_h + \tau_f$ and $\iota^* = 1 + \phi + \tau_f^* + \tau_h^*$. A key point is that, when a government selects its politically optimal export policy, it does so to deliver its preferred local price abroad for its domestically produced varieties, which in turn *neutralizes* the externality that travels from the trading partner's import tariff through this price. Likewise, a government's politically optimal import tariff delivers its preferred local price in the domestic market for varieties produced abroad, which in turn neutralizes the externality that travels from the trading partner's export policy through this price.

Thus, while the monopolistic competition model admits a rich set of local-price externalities that complement the traditional terms-of-trade externality, the local-price externalities are "shut down" when each government selects its politically optimal import and export policies, leaving only the terms-of-trade externality, which by itself amounts simply to a lump-sum transfer between governments. In this sense, the terms-of-trade externality remains the sole rationale for a trade agreement. From this perspective, it is also apparent that politically optimal trade policies would not in general be efficient were governments to possess an incomplete set of trade-policy instruments. For example, if export policies were unavailable, then a government would not be able to use its export policy to deliver its preferred local price abroad for domestically produced varieties, and its trading partner's import tariff would then induce a local-price externality through this channel.[ad] As in the partial-equilibrium perfect competition model discussed earlier, the

[ad] Indeed, in this model, import tariffs do not generate a terms-of-trade externality, and so the inefficiency that emerges under Nash policies when export policies are unavailable clearly does not derive from this externality.

efficiency of politically optimal policies relies deeply on the assumption that governments possess a complete set of trade-policy instruments.[ae]

The finding just described—that politically optimal policies are efficient in the monopolistic competition model when governments have a complete set of trade-policy instruments—extends to a range of other imperfect competition settings. For example, Bagwell and Staiger (2015) show that this finding holds as well in the Cournot delocation model considered by Venables (1985), wherein firms engage in Cournot competition and markets are segmented. In complementary work, Bagwell and Staiger (2012b) show that the finding holds as well in various "profit-shifting" models, where the number of firms is fixed and trade policy can shift profits from one country to another. An important direction for future research concerns the extent to which this finding extends to settings with multiple countries, domestic policies and other forms of imperfect competition.[af] Another extension, which we discuss below, concerns whether this finding extends in "offshoring" settings where prices may be determined through bilateral bargaining between sellers and buyers.

Finally, while the results summarized here indicate that the terms-of-trade externality remains the sole rationale for a trade agreement in important imperfect-competition settings, non-terms-of-trade externalities may nevertheless be important for understanding key features of actual trade agreements. First, actual trade agreements may constrain the trade-policy instruments that are available to governments, so that local-price externalities are not neutralized when politically optimal policies are selected. As one important example, we note that the WTO prohibits export subsidies. As Ossa (2011) argues and as

[ae] The argument here is distinct from standard arguments that governments may need a complete set of policy instruments in order to achieve a first-best outcome in the presence of market imperfections. The argument here instead concerns whether governments can achieve an efficient outcome while using politically optimal policies, when efficiency and political optimality are defined relative to a fixed set of policy instruments.

[af] In this regard, Campolmi et al. (2014) extend the study of trade agreements within a Venables (1987) delocation-type model to include domestic policies as well as trade policies. They argue in this setting that the choice of domestic policies introduces a novel motive for non-cooperative trade policy choices; and they further claim that politically optimal policies are not efficient in this setting, implying the possibility of a trade agreement whose purpose extends beyond the internalization of terms-of-trade externalities. However, the characterization of politically optimal policies employed by Campolmi et al. does not conform to the definition of such policies that we have described here—Campolmi et al. impose the restriction that governments act as if their unilateral policy choices had no impact on world prices, while as described earlier (see also Bagwell and Staiger, 2015) we impose the restriction that governments act as if they did not *value* the world price impacts of their unilateral policy choices; and they evaluate their political optimum conditions at reciprocal free trade policies, which do not correspond to the politically optimal policies in this setting according to our definition—and so we view the purpose of trade agreements in this setting as still an open question. See also Costinot et al. (2015), who consider a modeling framework that includes that of Campolmi et al. as a special case, and who argue that the motives for policy intervention can indeed be understood to reflect only terms-of-trade considerations in this setting.

we discuss further in a later section, in such a restricted-instrument setting, novel externalities may influence trade-agreement purpose and design. Second, and relatedly, models with imperfect competition may deliver novel interpretations of certain design features of trade agreements. For instance, as Bagwell and Staiger (2012a) show, the Cournot delocation model can provide a novel interpretation of the WTO's restrictions on export subsidies.[ag] Finally, our discussion here emphasizes international externalities that travel through prices. Trade policies also may be associated with international nonpecuniary externalities, such as global warming. The purpose and design of trade agreements in settings characterized by pecuniary and nonpecuniary international externalities is a very important direction for future research.[ah]

2.4 Offshoring Model of Trade Agreements

In all of the models that we describe earlier, prices are determined by market-clearing mechanisms. A growing volume of trade, however, involves intermediate inputs, with firms "offshoring" production and frequently customizing inputs to reflect the needs of buyers. In the presence of such "relationship-specific" investments, hold up is a natural concern. Since contracts involving international transactions may be difficult to enforce, the resulting prices may be determined by bilateral negotiations rather than market-clearing mechanisms: that is, while it is natural to think that market clearing is still a feature of the equilibrium in settings where offshoring is prevalent, the discipline that market-clearing places on the determination of international prices is likely to be diminished. An interesting issue is whether the rise of offshoring impacts in some fundamental way the purpose and design of trade agreements. Antràs and Staiger (2012a,b) examine this issue in detail. Here, we develop the basic model used by Antràs and Staiger (2012a), characterize the Nash, efficient and politically optimal policies, and emphasize in particular their finding that the politically optimal policies are not efficient when governments have political-economic motivations and seek to use trade policies for the purposes of redistribution. The key implication of this finding is that the inefficiency associated with the terms-of-trade externality is not necessarily the only problem for a trade agreement to solve when intermediate inputs are traded and the resulting prices are determined by bilateral negotiations.

[ag] See also Bagwell and Lee (2015) for a recent effort that interprets these restrictions from the perspective of the monopolistic competition model with heterogeneous firms developed by Melitz and Ottaviano (2008).

[ah] Related existing research that emphasizes the presence of international nonpecuniary externalities includes Limão (2005) and Spagnola (1999a,b). Related research on trade agreements in the presence of *domestic* nonpecuniary externalities includes Ederington (2001) and Lee (2007, 2014).

2.4.1 Offshoring Model

Following Antràs and Staiger (2012a), we consider two countries that are small in the market for a final good, which is called good 1 and is traded on world markets at a price of unity. Consumers in the home (H) and foreign (F) countries share identical preferences, which take the quasi-linear form

$$U^j = c_0^j + u(c_1^j), \tag{43}$$

where c_i^j is the consumption of good $i \in \{0, 1\}$ in country $j \in \{H, F\}$ and where $u' > 0 > u''$. As is standard, the numeraire good, which here is good 0, is costlessly traded and consumed in positive quantities in both countries. With p_1^j denoting the price of good 1 in country j, we let $D_1(p_1^j) \equiv u'^{-1}(p_1^j)$ indicate the demand for good 1 in country j. The corresponding consumer surplus function is then represented as

$$CS(p_1^j) = \int_{p_1^j}^{\bar{p}} D_1(p)dp, \tag{44}$$

where $\bar{p}$ is the choke price (if any).

Good 1 is produced by the home country using a customized input x sourced in the foreign country, where the good-1 production function $\gamma(x)$ satisfies $\gamma(0) = 0, \gamma'(x) > 0 > \gamma''(x)$, $\lim_{x\to 0}\gamma'(x) = \infty$ and $\lim_{x\to\infty}\gamma'(x) = 0$.[ai] The marginal cost for a foreign input supplier is normalized to unity.

The home country H has a unit mass of producers of the final good 1, while the foreign country F has a unit mass of suppliers of the intermediate input good x. Importantly, an input requires customization for its intended final good producer and is thus specific to the particular supplier–producer relationship. A simplifying assumption is that a given input in fact has no outside value to other suppliers. As well, Antràs and Staiger assume that contracts are infeasible so that the price at which each supplier in F sells its inputs to a producer in H is determined ex post (after the cost of producing x is sunk) according to the generalized Nash bargaining solution, where $\alpha \in (0, 1)$ is the weight attached to the home producer and $1 - \alpha \in (0, 1)$ is thus the weight attached to the foreign supplier.

Before supplier–producer relationships are established, trade policies are determined. Let τ_x^H denote the specific trade tax imposed by H on imports of x from F, where $\tau_x^H > 0$ ($\tau_x^H < 0$) indicates an import tariff (import subsidy). Similarly, τ_x^F denotes the specific trade tax imposed by F on exports of x to H, where $\tau_x^F > 0$ ($\tau_x^F < 0$) indicates an export tariff (export subsidy). It is convenient also to define the total trade tax on the intermediate input x as $\tau_x \equiv \tau_x^H + \tau_x^F$. The model allows that the home country H may import or export final good 1. Let the specific trade tax imposed by H on the final good

[ai] An important implication of concavity is that $\gamma(x)/x > \gamma'(x)$ for $x > 0$.

1 be represented as τ_1^H, where $\tau_1^H > 0$ ($\tau_1^H < 0$) indicates an import tariff or export subsidy (import subsidy or export tariff).[aj]

Formally, Antràs and Staiger (2012a) consider a game with the following stages:

stage 0: Trade policies τ_x^H, τ_x^F and τ_1^H are determined.

stage 1: Each supplier in F is randomly matched to a unique producer in H.

stage 2: Each supplier decides on an amount x of the customized input to produce.

stage 3: Each producer–supplier pair bargains over the price of the intermediate input.

stage 4: Each producer imports x units and produces the final good, and payments and trade taxes are settled.

To begin, let us consider the bargaining game in *stage* 3, at which point the trade policies are fixed, the volume x is determined, and the associated production costs for x are sunk. The joint surplus over which a producer–supplier pair bargains in *stage* 3 can be represented as

$$J(\tau_1^H, \tau_x, x) = (1 + \tau_1^H) y(x) - \tau_x x, \tag{45}$$

where $p_1^H = 1 + \tau_1^H \equiv p_1^H(\tau_1^H)$ is the price of final good 1 and we recall that $\tau_x \equiv \tau_x^H + \tau_x^F$ is the total trade tax on the intermediate good x. The producer in H and the supplier in F thus obtain the respective bargaining payoffs of $\alpha J(\tau_1^H, \tau_x, x)$ and $(1-\alpha) J(\tau_1^H, \tau_x, x)$.

Foreseeing this bargaining payoff, a supplier in F chooses the volume x in *stage* 2 so as to maximize $(1-\alpha) J(\tau_1^H, \tau_x, x) - x$, where we recall that the marginal production cost of x is unitary. Let us denote the profit maximizing volume as $\hat{x} = \hat{x}(\tau_1^H, \tau_x)$, where this volume is defined by the first-order condition for profit maximization:

$$(1-\alpha)(1 + \tau_1^H) y'(\hat{x}) = (1-\alpha)\tau_x + 1. \tag{46}$$

Given the maintained assumptions that $1 + \tau_1^H > 0$ and $y'' < 0$, it is straightforward to verify that $\partial \hat{x} / \partial \tau_1^H > 0 > \partial \hat{x} / \partial \tau_x$. Intuitively, a tariff change elicits a higher profit-maximizing value for x when the tariff change results in a higher joint surplus for bargaining.

It is instructive to pause at this point and highlight the hold-up problem that is embedded in the model. If the producer–supplier pair could contract over x with the objective of maximizing the sum of their joint payoffs inclusive of the cost of producing x, then they would choose x to maximize $J(\tau_1^H, \tau_x, x) - x$. The associated first-order condition is

[aj] As Antràs and Staiger (2012a) argue, the foreign country F has no incentive to deviate from free trade in the final good, and so for our discussion here we simply assume that F maintains free trade in the final good.

$$(1+\tau_1^H)\gamma'(x)=\tau_x+1. \tag{47}$$

Comparing (47) with (46), we conclude that $\hat{x}$ is lower than the level that would be contracted upon. The key point, of course, is that the cost that the foreign supplier incurs in producing the intermediate input is treated as sunk once the Nash bargaining process commences in *stage* 3.

Continuing with our analysis of the Antràs and Staiger (2012a) game, we now roll back to *stage* 1 to determine payoffs for the home producer and foreign supplier. We define these payoffs as follows:

$$\begin{aligned}\pi^H&=\alpha J(\tau_1^H,\tau_x,\hat{x}(\tau_1^H,\tau_x))\equiv\pi^H(\tau_1^H,\tau_x)\\ \pi^F&=(1-\alpha)J(\tau_1^H,\tau_x,\hat{x}(\tau_1^H,\tau_x))-\hat{x}(\tau_1^H,\tau_x)\equiv\pi^F(\tau_1^H,\tau_x).\end{aligned} \tag{48}$$

Our next step is to represent welfare functions for the governments of the home and foreign countries. Letting $\gamma^j\geq 1$ denote the welfare weight that the government of country j attaches to the payoffs enjoyed by its firms, the home- and foreign-country government welfare functions may be, respectively, represented as follows:

$$\begin{aligned}W^H(\tau_1^H,\tau_x^H,\tau_x^F)&=CS(1+\tau_1^H)+\gamma^H\pi^H(\tau_1^H,\tau_x)+\tau_1^H[D_1(1+\tau_1^H)-\gamma(\hat{x}(\cdot))]+\tau_x^H\hat{x}(\cdot)\\ W^F(\tau_1^H,\tau_x^H,\tau_x^F)&=CS(1)+\gamma^F\pi^F(\tau_1^H,\tau_x)+\tau_x^F\hat{x}(\cdot)\end{aligned} \tag{49}$$

where $\hat{x}(\cdot)\equiv\hat{x}(\tau_1^H,\tau_x)$. Notice that the foreign country imports final good 1 at the world price, whereas the home-country government may be tempted to use its final-good tariff τ_1^H so as to influence the joint bargaining surplus and thus the determination of $\hat{x}$. The final two terms in the top line of (49) and the final term in the bottom line of (49) are tariff-revenue terms.

We may now represent world welfare, W^w, as the sum of W^H and W^F. Formally:

$$\begin{aligned}W^w&=W^H(\tau_1^H,\tau_x^H,\tau_x^F)+W^F(\tau_1^H,\tau_x^H,\tau_x^F)\\ &=CS(1+\tau_1^H)+CS(1)+\gamma^H\pi^H(\tau_1^H,\tau_x)+\gamma^F\pi^F(\tau_1^H,\tau_x)\\ &\quad+\tau_1^H[D_1(1+\tau_1^H)-\gamma(\hat{x}(\cdot))]+\tau_x\hat{x}(\cdot)\\ &\equiv W^w(\tau_1^H,\tau_x).\end{aligned} \tag{50}$$

Observe that W^w depends on τ_x^H and τ_x^F only through their sum, τ_x.

Having described the basic model, we now show that the welfare functions can be written as functions of local and world prices. As in the models described earlier, such a formulation is useful, since it enables us to define politically optimal policies. We begin by defining the implied prices for the offshoring model.

We may think of a supplier in F as delivering $\hat{x}$ units to a producer in H at the home-country local price p_x^H, where p_x^H is thus defined by

$$p_x^H \hat{x} - (1 + \tau_x^H + \tau_x^F)\hat{x} \equiv \pi^F(\tau_1^H, \tau_x). \tag{51}$$

Using (45) and (48), we may now rewrite (51) equivalently as

$$p_x^H = \frac{(1-\alpha)(1+\tau_1^H)\gamma(\hat{x}(\cdot))}{\hat{x}(\cdot)} + \alpha\tau_x \equiv p_x^H(\tau_1^H, \tau_x). \tag{52}$$

Next, we let p_x^* represent the international or world price for the intermediate good x. This is the implied price that prevails prior to the application of H's import policy:

$$p_x^* = p_x^H - \tau_x^H = \frac{(1-\alpha)(1+\tau_1^H)\gamma(\hat{x}(\cdot))}{\hat{x}(\cdot)} + \alpha\tau_x^F - (1-\alpha)\tau_x^H \equiv p_x^*(\tau_1^H, \tau_x^H, \tau_x^F). \tag{53}$$

Finally, we may define the foreign-country local price p_x^F as the implied price that prevails before the application of F's export policy:

$$p_x^F = p_x^* - \tau_x^F = \frac{(1-\alpha)(1+\tau_1^H)\gamma(\hat{x}(\cdot))}{\hat{x}(\cdot)} - (1-\alpha)\tau_x \equiv p_x^F(\tau_1^H, \tau_x). \tag{54}$$

We now make a few observations. First, we observe from (53) and (54) that $p_x^H - p_x^F = \tau_x$. Second, we recall that $p_1^H = 1 + \tau_1^H$. Next, using these observations, we note that we may think of $\hat{x}(\tau_1^H, \tau_x)$ as a function of local prices. Formally, we may define the function $\bar{x}$ as follows:

$$\bar{x}(p_1^H, p_x^H - p_x^F) = \hat{x}(\tau_1^H, \tau_x). \tag{55}$$

Finally, with $\bar{x}$ defined in terms of prices as indicated in (55), we can likewise express firm payoffs and tariff revenues in terms of prices. For instance, using (51) and replacing $\hat{x}$ with $\bar{x}$ as allowed via (55) and $\tau_x^H + \tau_x^F \equiv \tau_x$ with $p_x^H - p_x^F$, π^F can be expressed as a function of p_1^H, p_x^H and p_x^F. While firm profit depends only on total tariffs, tariff revenue is also influenced by the world price. For example, H's tariff revenue on the intermediate good, $\tau_x^H \hat{x}$, can be written as $(p_x^H - p_x^*)\bar{x}(p_1^H, p_x^H - p_x^F)$.

Proceeding in this general manner, we may now represent the home and foreign government welfare functions, and thus the world welfare function, in terms of the local and world prices that the underlying trade policies imply. We begin with the welfare function of H's government:

$$\begin{aligned} W^H &= CS(p_1^H) + \gamma^H[p_1^H\gamma(\bar{x}(\cdot)) - p_x^H\bar{x}(\cdot)] + (p_1^H - 1)[D(p_1^H) - \gamma(\bar{x}(\cdot))] + (p_x^H - p_x^*)\bar{x}(\cdot) \\ &\equiv \overline{W}^H(p_1^H(\tau_1^H), p_x^H(\tau_1^H, \tau_x), p_x^F(\tau_1^H, \tau_x), p_x^*(\tau_1^H, \tau_x^H, \tau_x^F)), \end{aligned} \tag{56}$$

where $\bar{x}(\cdot) \equiv \bar{x}(p_1^H, p_x^H - p_x^F)$. And next, the welfare function of F's government:

$$\begin{aligned} W^F &= CS(1) + \gamma^F[p_x^F - 1]\bar{x}(\cdot) + (p_x^* - p_x^F)\bar{x}(\cdot) \\ &\equiv \overline{W}^F(p_1^H(\tau_1^H), p_x^H(\tau_1^H, \tau_x), p_x^F(\tau_1^H, \tau_x), p_x^*(\tau_1^H, \tau_x^H, \tau_x^F)). \end{aligned} \tag{57}$$

Finally, the world welfare function can now be defined as follows:

$$\begin{aligned} \overline{W}^w &= \overline{W}^H + \overline{W}^F = CS(p_1^H) + \gamma^H[p_1^H \gamma(\bar{x}(\cdot)) - p_x^H \bar{x}(\cdot)] \\ &\quad + (p_1^H - 1)[D(p_1^H) - \gamma(\bar{x}(\cdot))] + CS(1) + \gamma^F[p_x^F - 1]\bar{x}(\cdot) \\ &\quad + (p_x^H - p_x^F)\bar{x}(\cdot) \\ &\equiv \overline{W}^w(p_1^H(\tau_1^H), p_x^H(\tau_1^H, \tau_x), p_x^F(\tau_1^H, \tau_x)). \end{aligned} \tag{58}$$

We note that the welfare functions of the governments of H and F each depend on the terms of trade, p_x^*. The government of H (F) enjoys a terms-of-trade gain when p_x^* is lower (higher):

$$\overline{W}^H_{p_x^*} = -\bar{x}(\cdot) < 0 \text{ and } \overline{W}^F_{p_x^*} = \bar{x}(\cdot) > 0. \tag{59}$$

But the world welfare function is independent of the terms of trade: for given local prices, and thus for a given value of $\bar{x}(\cdot)$, a change in p_x^* simply amounts to an international transfer.

2.4.2 Prisoners' Dilemma

With the offshoring model developed and the welfare functions presented in terms of local and world prices, we are prepared now to characterize the Nash, efficient and politically optimal trade policies. To this end, we first present the associated first-order conditions and then describe the main findings. For each optimization problem, we assume that the corresponding second-order conditions are satisfied.

The *Nash trade policies* for the offshoring model, $(\tau_1^{HN}, \tau_x^{HN}, \tau_x^{FN})$, satisfy the following first-order conditions:

$$\begin{aligned} \overline{W}^H_{p_1^H} + \overline{W}^H_{p_x^H}\frac{\partial p_x^H}{\tau_1^H} + \overline{W}^H_{p_x^F}\frac{\partial p_x^F}{\tau_1^H} - \bar{x}(\cdot)\frac{\partial p_x^*}{\tau_1^H} = 0 \\ \overline{W}^H_{p_x^H}\frac{\partial p_x^H}{\tau_x} + \overline{W}^H_{p_x^F}\frac{\partial p_x^F}{\tau_x} - \bar{x}(\cdot)\frac{\partial p_x^*}{\tau_x^H} = 0 \\ \overline{W}^F_{p_x^H}\frac{\partial p_x^H}{\tau_x} + \overline{W}^F_{p_x^F}\frac{\partial p_x^F}{\tau_x} + \bar{x}(\cdot)\frac{\partial p_x^*}{\tau_x^F} = 0, \end{aligned} \tag{60}$$

where we use $p_1^H = 1 + \tau_1^H$, $\tau_x = \tau_x^H + \tau_x^F$ and (59). Next, the *efficient trade policies* for the offshoring model, (τ_1^{HE}, τ_x^E), satisfy the following first-order conditions:

$$\begin{aligned}\overline{W}^{w}_{p_1^H} + \overline{W}^{w}_{p_x^H}\frac{\partial p_x^H}{\tau_1^H} + \overline{W}^{w}_{p_x^F}\frac{\partial p_x^F}{\tau_1^H} = 0\\ \overline{W}^{w}_{p_x^H}\frac{\partial p_x^H}{\tau_x} + \overline{W}^{w}_{p_x^F}\frac{\partial p_x^F}{\tau_x} = 0.\end{aligned} \tag{61}$$

Finally, the *politically optimal trade policies* for the offshoring model, $(\tau_1^{HPO}, \tau_x^{HPO}, \tau_x^{FPO})$, satisfy the following first-order conditions:

$$\begin{aligned}\overline{W}^{H}_{p_1^H} + \overline{W}^{H}_{p_x^H}\frac{\partial p_x^H}{\tau_1^H} + \overline{W}^{H}_{p_x^F}\frac{\partial p_x^F}{\tau_1^H} = 0\\ \overline{W}^{H}_{p_x^H}\frac{\partial p_x^H}{\tau_x} + \overline{W}^{H}_{p_x^F}\frac{\partial p_x^F}{\tau_x} = 0\\ \overline{W}^{F}_{p_x^H}\frac{\partial p_x^H}{\tau_x} + \overline{W}^{F}_{p_x^F}\frac{\partial p_x^F}{\tau_x} = 0.\end{aligned} \tag{62}$$

We now consider the efficiency of the Nash and politically optimal policies. Looking at (60) and (61), it is natural to expect that the Nash trade policies are inefficient. After all, when setting their Nash policies, governments are mindful of the impact of their policies on the terms of trade, p_x^*, even though for fixed local prices the terms of trade has no impact on world welfare. Antràs and Staiger (2012a) show that the Nash policies are indeed inefficient, and thus that a problem exists for a trade agreement to solve.

But is the terms-of-trade externality the only problem for a trade agreement to address in this setting? Antràs and Staiger show that politically optimal trade policies are efficient when $\gamma^F = 1$, so that the foreign-country government maximizes national income; however, they find that, if $\gamma^F > 1$, then politically optimal trade policies are inefficient. Thus, in the offshoring model and when governments have political-economic motivations and value redistribution, it follows that the terms-of-trade externality is not the only problem for a trade agreement to address.

At a broad level, why might politically optimal policies be inefficient in the offshoring model?[ak] Notice that the bottom two conditions in (62) can be added to deliver the bottom condition in (61); thus, the underlying source of the potential inefficiency of politically optimal policies is connected to the top conditions in (61) and (62) and thus to the determination of H's final-good trade tax, τ_1^H. Let us thus follow Antràs and Staiger (2012a) and contemplate a small increase in τ_1^H that is coupled with a change in τ_x^H that leaves the world price p_x^* fixed. This policy adjustment results in a higher value for p_x^H but leaves p_x^F fixed (since τ_x^F is unaltered). Starting at the politically optimal policies, Antràs and Staiger show that the resulting changes in p_1^H and p_x^H lead only to a second-order loss

[ak] Antràs and Staiger (2012a) also provide a more detailed explanation that clarifies the role played by $\gamma^F > 1$ in the inefficiency of the political optimum in the offshoring model.

for H's government welfare but can generate a first-order effect for the government welfare of F. Intuitively, F's government does not have an "offsetting" instrument with which to position $p_1^H = 1 + \tau_1^H$ at its preferred level and thereby "shut down" the associated local-price externality; hence, F's government may experience a first-order welfare change at politically optimal policies when H's government alters p_1^H by changing τ_1^H.[al] In fact, as Antràs and Staiger show, the described policy adjustment generates a first-order gain for F's government welfare when $\gamma^F > 1$. Politically optimal policies are thus inefficient, since τ_x^H and τ_x^F can then be adjusted while holding τ_x fixed so as to effect a transfer (via the resulting world-price change) that compensates H's government for its second-order welfare loss while still delivering a first-order welfare gain to F's government.

Our preceding discussion provides answers to the first and third questions that motivate this section. Specifically, for the offshoring model, Nash tariffs are inefficient, and if $\gamma^F > 1$ an inefficiency remains even when governments are not motivated by the terms-of-trade implications of their trade policies. We have not addressed for this model the second motivating question about the role of reciprocity. We postpone our discussion of reciprocity in the offshoring model until Section 3.3.

2.5 Summary

At this point, we have reviewed four models of trade agreements, and each model features a terms-of-trade externality.[am] We have also argued that the first three models can be interpreted as indicating that the sole purpose of a trade agreement is to help governments escape from a terms-of-trade driven Prisoners' Dilemma, at least when governments possess a complete set of trade policy instruments, in the particular sense that politically optimal policies are efficient in these models. We may therefore regard these first three models as falling within the "class" of terms-of-trade theories of trade agreements.

Are there some general features of these models that can be used to identify models that fall within the terms-of-trade class? Maggi (2014) makes some progress in providing

[al] Antràs and Staiger (2012a) work within a benchmark model where the offshoring of inputs occurs between H and F who are both small countries on world markets for the final good. It might be thought that this small country assumption is responsible for the inefficiency of the political optimum in this setting, because it generates a "missing instrument" problem for F's government with respect to the final good price in H's market. However, in their Online appendix G Antràs and Staiger confirm the inefficiency of the political optimum in a three-large-country extension of their benchmark model, while in their Online appendix H they show that when international prices are determined in their benchmark model by standard market-clearing conditions rather than bilateral bargaining the political optimum is indeed efficient. Together these results indicate that the missing-instrument problem described in the text is associated with the nature of international price determination rather than the small-country assumption.

[am] Below we will also discuss the commitment theory of trade agreements, which does not feature a terms-of-trade externality. We do so in the context of our evaluation of reciprocity as a design feature of trade agreements.

an answer to this question. He identifies three conditions that together are sufficient for efficiency of the political optimum: (i) there are only two countries; (ii) there are no income effects; and (iii) governments choose only (a complete set of) trade taxes. Indeed, each of the first three models we have reviewed earlier satisfies the sufficient conditions that Maggi identifies.[an] However, it is important to realize that this set of sufficient conditions is by no means necessary. On the contrary, the literature includes a number of findings that the political optimum is efficient also when these sufficient conditions are not met, and specifically in general equilibrium models with income effects (see for example Bagwell and Staiger, 1999 and DeRemer, 2012, appendix E), in models featuring domestic policy instruments in addition to a complete set of trade policy instruments (see for example Bagwell and Staiger, 2001a, 2002; Staiger and Sykes, 2011, and DeRemer, 2013), and in models with more than two countries (see for example Bagwell and Staiger, 1999, 2001b).

We emphasize two take-away points from our discussion just above. First, it is not a simple matter to generate models of trade agreements that fall outside the terms-of-trade class (beyond the commitment theory of trade agreements which we discuss below). We have described here two ways that such models have been generated: missing instruments, and international prices determined by bilateral bargaining.[ao] There may well be other possibilities, but thus far they have not been identified in the literature. And second, while it is useful to identify sets of sufficient conditions for models to fall within the terms-of-trade theory class, working through the details of models that may satisfy these sufficient conditions is nevertheless illuminating for gaining a deeper understanding of the features that ultimately dictate whether or not escape from a terms-of-trade driven Prisoners' Dilemma can be said to be the sole purpose of a trade agreement.

2.6 GATT's Designers and the Terms-of-Trade Externality

As reflected in the previous sections, the terms-of-trade theory provides an important benchmark in the literature for interpreting and evaluating the design of trade agreements. And as we have indicated, the design of a trade agreement is likely to reflect its purpose. In this light, it is important to ask whether there is meaningful contact between the problems emphasized by this theory and those emphasized by the designers of the institution it is supposed to illuminate. If the problems to be addressed as perceived by those involved in the design of a trade agreement make little or no substantive contact with the problems that the terms-of-trade theory suggests should be at the forefront of their thinking, then findings that the design features of the agreement should work well or poorly to solve the problems at the center of the theory are less meaningful,

[an] And as Maggi (2014) emphasizes, surprisingly, politically optimal policies can be efficient even in models that feature nonpecuniary international externalities as long as these sufficient conditions are satisfied.

[ao] A third way, which we discuss below, arises when there are many countries and MFN is not imposed.

and the position that the agreement's successes or failings can be understood on these terms is less tenable.

Before moving on with our survey, we therefore pause here and ask: Did the designers of GATT, or at least economists at the time including those directly involved in GATT's design, emphasize the terms-of-trade externality associated with commercial policy? And if so, did they view GATT as a forum for addressing these terms-of-trade externalities?

Some suggestive evidence on these questions is contained in what became known as the *Haberler Report*, commissioned by GATT and written by a Panel of Experts composed of Roberto de Oliveira Campos, Gottfried Haberler, James Meade and Jan Tinbergen.[ap] The Panel's terms of reference were to investigate the prevalence of agricultural protectionism and "...the failure of the export trade of the under-developed countries to expand at a rate commensurate with their growing import needs." (GATT, 1958, foreword).

The Haberler Report provides a wide-ranging discussion of the economic issues of the day and their impacts on the exports of developing countries, emphasizing business cycle fluctuations and balance of payments constraints as well as commercial policy concerns. But when it comes to commercial policy and the topic of agricultural protectionism, terms-of-trade externalities, albeit expressed not in the simple two-country context that we have considered above but rather in the context of a more complex multicountry setting along the lines that we introduce in a later section of this chapter, appears to be at the center of the Report's discussion:

> *The problem of the interests of different primary producing countries outside industrialized Western Europe and North America is ... not only a question which of the other countries would gain by a moderation of agricultural protectionism in these two great industrialized regions; there are undoubtedly cases in which an increase in agricultural protectionism in these two regions, while it would be to the disadvantage of some of the unindustrialized countries, would actually be to the advantage of others. Two examples will serve to illustrate the point. An increased stimulus to the production of wheat in any of the countries of North America or of Western Europe by increasing the exportable surplus of North America and decreasing the import requirements of Western Europe would depress the world market for wheat. This might mean that a country like India or Japan would obtain cheaper imports of wheat (either because of a fall in the world price or because of a development of special sales or gifts for the disposal of surplus wheat by the United States), but a country like Australia or the Argentine which competed in the world export market for wheat would be damaged. Another example of the same principle would be provided by measures which stimulated the export of raw cotton from the United States: this might increase the plenty and cheapness of raw cotton in world markets; an importing country like Japan would gain but competing exporters like Egypt, the Sudan, and Brazil would lose.*

[ap] James Meade was a member of the British delegation to the London and Geneva conferences in 1946 and 1947 which produced the charter for the International Trade Organization and GATT. Along with Keynes, Meade was widely regarded as a central figure in these conferences (see for example Penrose, 1953, pp. 89–90).

> *In general, if one considers any particular agricultural product, a protective stimulus to its production in any one country by increasing supplies relatively to the demand for that product will tend to depress the world market for that product. This will damage the interests of other countries which are exporters of the product on the world market. But it will be to the national interest of countries which import the product from world markets. Whether the initial protective stimulus confers a net benefit or a net damage to all other countries concerned depends, therefore, upon whether the country giving the protective stimulus to its own production is an exporter or an importer of the product; if it is an exporter it is conferring a benefit on the world by giving its supplies away at a cheap price; if it is an importer it is damaging the rest of the world by refusing to take their supplies.*
>
> *This general principle can be applied to a single country or to a whole region. It is because Western Europe and North America in combination are net* importers *of agricultural produce that we reach the general conclusion that a reduction of agricultural protectionism in these areas will on balance benefit the rest of the world...(GATT, 1958, pp. 93–94, original emphasis, footnotes omitted).*

In describing the impacts of agricultural protectionism in Western Europe and North America on various countries in the rest of the world, the Report's references to "depress the world market," "fall in the world price," gains for other importing countries from "cheaper imports" and losses for countries who are "competing exporters" hew quite closely to a terms-of-trade logic. And the references to the protective policies of "any of the countries of North America or of Western Europe," "any one country" and "a single country" suggest that the Report's authors did not find implausible the notion that a single country's protective choices could have world price impacts; indeed, the general principle for signing the international externalities associated with commercial policy intervention is couched in terms of "the country giving the protective stimulus." Hence, in these paragraphs the authors of the Haberler Report appear to be describing the terms-of-trade externality that is at the heart of the terms-of-trade theory of trade agreements.

Moreover, the Haberler Report makes recommendations to GATT that are based on these terms-of-trade externality patterns. For example, the following recommendation, taken from the Report's executive summary, reflects the application of the international externality signing principle as articulated in the quoted excerpt above:

> *Since in North America and Western Europe as a whole net imports of agricultural products represent the relatively narrow margin by which their large domestic consumption exceeds their large, but not quite so large, domestic production, a relatively small restraint on domestic production or stimulus to domestic consumption could lead to a large percentage increase in their net imports. For this reason much could be achieved by some moderate change in the direction of the agricultural policies of the highly industrialized countries. (GATT, 1958, p. 9).*

There are many other such examples throughout the commercial policy portion of the Report. It appears as well, then, that the authors of the Haberler Report viewed GATT as a forum for addressing these terms-of-trade commercial policy externalities.

This, of course, does not establish that GATT is well designed to solve the terms-of-trade driven Prisoners' Dilemma, or even necessarily that there is a terms-of-trade driven

Prisoners' Dilemma to be solved: though implicit in its recommendations, the Haberler Report is silent on its position about a key step in the terms-of-trade logic, that in not valuing the terms-of-trade externalities imposed by their commercial policy choices governments were led to make unilateral commercial policy choices that were overly protective from an international efficiency perspective. And in any event these quotes are not a substitute for systematic empirical evidence (some of which we survey below) relating to the terms-of-trade theory's relevance for interpreting GATT/WTO outcomes. But in suggesting that the designers of GATT placed emphasis on the terms-of-trade externalities associated with the commercial policy choices of individual countries, and that they viewed GATT as a forum for addressing these externalities, the statements quoted above lend credence to the view that GATT *could* be especially well designed to solve the terms-of-trade problem—in part perhaps as a result of experimentation and purposeful engineering and in part perhaps by luck—that the GATT/WTO may in large part owe its successes to these design features, and that its failings might also be understood at least in part from the perspective of this theory.

A final observation is also relevant: while we have presented the terms-of-trade theory using the "world price" and "terms-of-trade" language that economists typically employ to describe the relevant policy externalities, the theory can easily be translated into the language of "market access" that real-world trade-policy negotiators prefer. When a country raises its import tariff and thereby shifts in its import demand curve, the resulting price effect under which it enjoys a terms-of-trade improvement is accompanied by a volume effect under which its trading partner experiences a reduction in access to its market. Once this link between price and volume effects is forged, the terms-of-trade theory can be reexpressed in the market access language that trade-policy negotiators adopt.[aq]

We now turn to the literature on the key design features of trade agreements. We emphasize three prominent features of the GATT/WTO: the principle of reciprocity, the principle of nondiscrimination as embodied in MFN, and tariff caps that allow for "binding overhang."

3. RECIPROCITY

In this section we consider the GATT principle of reciprocity. We begin with a discussion of reciprocal liberalization, define the principle of reciprocity in GATT, describe its applications in the GATT/WTO, consider its implications for sustainable bargaining outcomes, and consider its impacts on the GATT tariff bargaining process. Our initial discussion of the implications of reciprocity adopts the perspective of the terms-of-trade theory. We then consider as well the implications of reciprocity when a number of the

[aq] Bagwell and Staiger (2002) provide a formal definition of market access and further development of the relationship between the terms-of-trade theory and the language of market access.

standard assumptions which typically accompany analyses based on the terms-of-trade theory are relaxed. Finally, we close this section with a discussion of reciprocity from modeling perspectives that fall outside the terms-of-trade theory.

3.1 Reciprocal Liberalization

At a general level, reciprocity in trade agreements can be thought of as a norm or rule stating that negotiated tariff changes should result in tariff movements in the same direction across the participating countries. When negotiations are aimed at liberalizing market access, as is the focus of multilateral GATT/WTO negotiating rounds under GATT Article XXVIII bis, the reciprocity norm is that the tariffs of each negotiating partner should be reduced. And if one country increases its tariff, as in the context of *re*-negotiations under GATT Article XXVIII or the settlement of disputes about negotiated market access commitments under GATT Article XXIII, then the tariffs of its trading partners should rise as well. The expectation that reciprocity in market access commitments will be achieved and maintained in the GATT/WTO is fundamental to the institution.[ar]

But the reciprocity principle in GATT is more specific than simply a general complementarity in the direction of tariff changes across countries: it refers to a balance of tariff changes (in either a downward direction or an upward direction) that leads to changes in the volume of a country's exports that are commensurate with the changes in the volume of its imports.[as] In the context of negotiations over market access, therefore, GATT's principle of reciprocity is a negotiation norm that defines an idealized *terms of exchange* of market access, and it sets this terms of exchange at one for one: if country A wishes to acquire for its exporters a certain amount of additional access to the markets of country B, then under the GATT reciprocity norm it is expected that country A will in exchange provide to country B's exporters the same amount of additional access to its own markets; and the same terms of exchange applies to country B if country B seeks to obtain additional access for its exporters to the markets of country A.[at]

[ar] Indeed, GATT's reciprocity principle was thought to be critical to ensuring the constitutionality of US participation in GATT. See United States Council of the International Chamber of Commerce (1955, pp. 74–76).

[as] For example, in calibrating the magnitude of the Canadian tariff increase aimed at reducing exports from the EU that would be consistent with reciprocity and therefore permissible in response to the EU ban on imports of hormone-treated beef in the EC-Hormones dispute, the arbitrators (WTO, 1999, paragraph 41) stated: "To do so..., we have to focus on trade flows. We must estimate trade foregone due to the ban's continuing existence...".

[at] As we discuss in detail below, the principle of reciprocity arises as a negotiation norm when governments negotiate tariff reductions under GATT Article XXVIII bis, and the principle of reciprocity arises in GATT rules when tariffs are increased under GATT Article XXVIII renegotiations. See also Bagwell and Staiger (1999, 2002).

Two questions naturally follow. A first question is: Why did governments adopt the particular (one for one) terms of exchange embodied in the reciprocity principle of GATT? And a second question is: What advantage would there be for governments in choosing to fix the terms of exchange in GATT market access bargains in the first place?

3.1.1 The GATT Principle of Reciprocity

Why did this particular notion of reciprocity take hold in GATT? Put differently, what accounts for the feature that in the GATT/WTO the "price" of "purchasing" one additional unit of export market access is set equal to one additional unit of import market access?[au] Even if the price is to be fixed, why not a different price or terms of exchange, such as two additional units of exports for one additional unit of imports, or one additional unit of exports for three additional units of imports? In fact, there is an immediate and simple answer to this question, as long as governments are committed to adopting a *common* terms of exchange applied uniformly across all countries: the adding-up constraint imposed by market clearing makes any other terms of exchange infeasible. Of course, this is just the observation that one country's exports are another's imports, and hence together all countries cannot increase their exports more (or less) than they increase their imports.

Formally, this point can be confirmed very simply in any model that exhibits market clearing as a feature of the equilibrium outcome. Here we illustrate the point in a two-good two-country general equilibrium setting, remaining agnostic for now about the other features of the model economy (eg, the nature of competition or of international price determination).[av] For the purposes of defining reciprocity, we use the equilibrium terms of trade at original tariffs, $\widetilde{P}_0^w$, to convert "apples to oranges."[aw] With "Δ" in front of a variable denoting the change in that variable induced by a change in tariffs, let us begin with the home country, and consider a general version of reciprocity defined as

[au] As is well known, much of the design of GATT was inspired by the US Reciprocal Trade Agreements Act of 1934, and this included the particular notion of reciprocity adopted in GATT (see, for example, the discussion of this point in Penrose, 1953, p. 93), though as we discuss later in our chapter GATT allowed for an important multilateralization of the reciprocity principle. So although we pose our questions here in the context of GATT, they should be interpreted more broadly to include GATT's antecedents.

[av] In the Online Appendix (http://dx.doi.org/10.1016/bs.hescop.2016.04.005), we show how the points we emphasize below regarding GATT's principle of reciprocity and its implications within GATT/WTO practice extend to a many-good general equilibrium setting.

[aw] As long as countries agree to use the same conversion factor of apples to oranges in their assessments of reciprocity, for our purposes here the conversion factor can be anything they want, but the original world prices are a natural choice. Using world prices (original or new) as the conversion factor is important for the fixed-terms-of-trade property that we highlight next.

any change in tariffs that leads to a change in home-country export and import volumes satisfying

$$\Delta E = \gamma \tilde{P}_0^w \Delta M \tag{63}$$

where E and M denote home-country export and import volumes, and γ is a parameter specifying the terms of exchange of market access. Next observe that market clearing implies

$$\Delta E = \Delta M^* \tag{64}$$

$$\Delta E^* = \Delta M \tag{65}$$

where E^* and M^* denote foreign-country export and import volumes. But (63) and (65) imply $\Delta E = \gamma \tilde{P}_0^w \Delta E^*$, which using (64) then implies $\Delta M^* = \gamma \tilde{P}_0^w \Delta E^*$ or

$$\tilde{P}_0^w \Delta E^* = \frac{1}{\gamma} \Delta M^*, \tag{66}$$

which describes the foreign-country terms of exchange of market access that must accompany (63) according to the market-clearing requirements. Evidently, if a common terms of exchange is to be applied to both the home and the foreign country, then (63) and (66) imply $\gamma = 1$, which is to say a one-for-one exchange of import volumes for export volumes. Hence, if governments wish to adopt a common terms of exchange for all countries, the adding-up constraint imposed by market clearing makes it inevitable that they must adopt the one-for-one terms of exchange that characterizes GATT's reciprocity principle.

The essence of the first question posed above, then, is not why GATT's reciprocity principle reflects a one-for-one exchange, but rather why a common terms of exchange was adopted for all countries.[ax] This is a question that to our knowledge has not received specific attention in the trade agreements literature, but at a general level it seems plausible that part of the explanation may reflect a desire for fairness: there are reasons to think that fairness might originate as a social norm in a wide variety of bargaining settings (see, for example, Binmore, 2014), and ensuring a common terms of exchange for each country participating in GATT market access bargains resonates with a norm of fairness. In the context of answering the second question posed above, we will also suggest that the link between a common terms of exchange and fairness can be given a more specific representation in the trade agreements context, and as well that there may be additional efficiency benefits to adopting a common terms of exchange, provided that governments

[ax] An important exception to the reciprocity norm was granted to developing countries in the GATT/WTO. We touch on some of the implications of this "special and differential treatment" exception below (see Ornelas, 2016 for a comprehensive treatment).

seek (at least in part) to solve the terms-of-trade driven Prisoners' Dilemma problem described in Section 2.[ay]

We now turn to the second question: What advantage would there be for governments in choosing to fix the terms of exchange in GATT market access bargains in the first place? As we next describe, a number of potential benefits can be appreciated, once it is observed that negotiations that achieve the balance described by the GATT reciprocity principle leave the terms of trade unchanged. In this way, GATT's principle of reciprocity helps to create a bargaining forum within which, for each government, terms-of-trade manipulation is effectively removed from the calculus of preferred tariffs.[az]

The terms-of-trade fixing property of reciprocity can also be shown simply in any model that exhibits market clearing as a feature of the equilibrium outcome, provided that countries also satisfy a balanced trade condition. In fact, it is easy to show that reciprocity fixes the terms of trade even in the presence of nonzero trade imbalances, provided only that the size of the trade imbalances are not impacted by the tariff changes, but for simplicity we adopt here the assumption that trade is balanced.[ba] Staying within the two-good two-country general equilibrium setting described just above, and with reciprocity now defined by any set of tariff changes that satisfies

$$\Delta E = \tilde{P}_0^w \Delta M \tag{67}$$

$$\tilde{P}_0^w \Delta E^* = \Delta M^*, \tag{68}$$

for our present purposes we may focus on the home country, and we next introduce the home country's balanced trade condition. The balanced trade condition must hold both at the original and at new tariffs (as before, $\tilde{P}_0^w$ denotes the equilibrium terms of trade at original tariffs, and we now denote the equilibrium terms of trade at the new tariffs by $\tilde{P}_1^w$):

$$E_0 = \tilde{P}_0^w M_0 \text{ and } E_1 = \tilde{P}_1^w M_1. \tag{69}$$

Using $\Delta E \equiv E_1 - E_0$ and $\Delta M \equiv M_1 - M_0$, it then follows from (69) that (67) may be rewritten as

[ay] We have in mind that such benefits could help explain why a particular design feature might have been included in a trade agreement, either because the specific benefits suggested by the theory were understood to flow from the feature in question, or because the feature was incorporated by chance in some earlier agreement that performed well as a result of the benefits suggested by the theory.

[az] Bagwell and Staiger (1999) derive this property of reciprocity and highlight the benefits that we discuss below.

[ba] What is required is that the size of the new trade imbalance, measured at the new equilibrium world prices, must be the same as the size of the original trade imbalance, measured at original equilibrium world prices.

$$[\widetilde{P}_1^w - \widetilde{P}_0^w]M_1 = 0. \tag{70}$$

Hence, according to (70) and assuming only that a strictly positive volume of trade takes place at the new tariffs so that $M_1 > 0$, reciprocity exhibits a striking property: mutual changes in trade policy conform to the principle of reciprocity if and only if they leave the terms of trade unchanged.

We thus arrive at the following general conclusion. If governments wanted to create a forum for the exchange of market access commitments in which the terms of exchange were fixed at a common level for all countries, then they would have had to fix the terms of exchange of export market access for import market access at one for one, the same terms of exchange described by GATT's reciprocity principle. And having enshrined into the GATT reciprocity principle the only common terms of exchange available to them and provided that their trade was balanced (or if unbalanced, provided that the magnitude of the imbalances were independent of the outcome of tariff negotiations), governments would have, with GATT's reciprocity principle, directed the focus of GATT market access negotiations toward the volumes of trade desired by the participating governments rather than the terms of trade.

3.1.2 The Applications of Reciprocity in the GATT/WTO

Armed with this general conclusion, we consider next the specific applications of reciprocity within GATT/WTO practice. We highlight the potential implications of reciprocity with regard to addressing the terms-of-trade driven Prisoners' Dilemma problem; and for this purpose we now return to our benchmark two-good two-country perfectly competitive general equilibrium trade model described in Section 2. We first express our formal definition of reciprocity in terms of the notation introduced in that model. From an initial pair of tariffs, (τ^0, τ^{*0}), suppose that a tariff negotiation results in a change to the new pair of tariffs, (τ^1, τ^{*1}). Denoting the initial world and home local prices as $\widetilde{p}^{w0} \equiv \widetilde{p}^w(\tau^0, \tau^{*0})$ and $p^0 \equiv p(\tau^0, \widetilde{p}^{w0})$, and the new world and home local prices as $\widetilde{p}^{w1} \equiv \widetilde{p}^w(\tau^1, \tau^{*1})$ and $p^1 \equiv p(\tau^1, \widetilde{p}^{w1})$, we say that the tariff changes conform to *the principle of reciprocity* provided that

$$\widetilde{p}^{w0}[M(p^1, \widetilde{p}^{w1}) - M(p^0, \widetilde{p}^{w0})] = [E(p^1, \widetilde{p}^{w1}) - E(p^0, \widetilde{p}^{w0})], \tag{71}$$

where changes in trade volumes are valued at the existing world price.[bb] We next use the balanced trade condition (1)—which must hold at both the initial tariffs and the new tariffs—to confirm that (71) may be rewritten as

[bb] We have defined reciprocity here only for the domestic country, but as should now be clear in our two-country setting tariff changes conform to reciprocity for the domestic country if and only if they conform to reciprocity for the foreign country as well.

$$[\widetilde{p}^{w1} - \widetilde{p}^{w0}]M(p^1, \widetilde{p}^{w1}) = 0. \tag{72}$$

As (72) reflects and as stated above, mutual changes in trade policy conform to the principle of reciprocity if and only if they leave the terms of trade unchanged. We are now ready to consider the specific applications of reciprocity within GATT/WTO practice.

A first application of reciprocity can be found when negotiations are aimed at liberalizing market access, as is the focus of multilateral GATT/WTO negotiating rounds under GATT Article XXVIII bis. Suppose that governments begin with Nash tariffs. A key observation follows from a property of the Nash point that we emphasized earlier, namely, that using (4) and (5) the Nash first-order conditions (6) imply that $W_p < 0 < W_{p^*}^*$. The structure of international cost-shifting therefore implies that, beginning from their Nash tariff choices, each government would desire greater trade volume if this could be achieved at a fixed terms of trade. But if governments were to reduce tariffs according to reciprocity, then by (72) the terms of trade *would* be fixed, while the home local price p would fall and the foreign local price p^* would rise, allowing each government to achieve greater trade volume. Hence, as long as their tariff cuts are not too large, both the home-government welfare and the foreign-government welfare would then rise. In other words, starting at the Nash equilibrium and for tariff cuts that are not too large, liberalization under the principle of reciprocity is sufficient for mutual gains.[bc] Evidently, by directing the focus of GATT market access negotiations toward the volumes of trade desired by the participating governments at a fixed terms of trade, the principle of reciprocity provides a recipe for efficiency-enhancing gains from tariff liberalization.

A second application of reciprocity in the GATT/WTO can also be identified, in this case applying to situations where protective measures are being reimposed rather than liberalized. Such situations can arise in the GATT/WTO, both in the context of dispute resolution and in the context of the renegotiation of GATT tariff bindings. We focus here on renegotiation.[bd] The rules for renegotiation of GATT tariff bindings,

[bc] This sufficiency finding for the principle of reciprocity may be contrasted with our discussion in Section 2, where we show for various models that a general form of reciprocity is necessary for mutual gains.

[bd] Our focus is on the implications of reciprocity and whether it can be interpreted as serving a useful purpose rather than on its optimality per se. Maggi and Staiger (2015a) explore the optimal design of trade agreements in the presence of renegotiation more generally, and argue that reciprocity exhibits features that under certain conditions can be part of an optimal compensation rule in the event of disagreement. A related literature (eg, Lawrence, 2003; Bagwell, 2008; Beshkar, 2010; Grossman and Sykes, 2010; Maggi and Staiger, 2015b; and Staiger and Sykes, forthcoming) evaluates reciprocity in the context of GATT/WTO dispute resolution. As Mavroidis (2016) emphasizes, a distinguishing feature of reciprocity as it arises under dispute settlement is that the associated retaliation is prospective in nature and is available to the complainant only after the judicial process has run its course.

contained in GATT Article XXVIII, apply when one country reopens negotiations with its trading partners on a tariff binding to which it had earlier agreed, for the purpose of modifying (in an upward direction) or withdrawing the original tariff binding. These rules explicitly provide for the possibility that agreement might not be reached; and when this possibility arises, the country is permitted to modify or withdraw its original tariff binding anyway, with the understanding that the trading partners may then reciprocate. Here, the principle of reciprocity puts a lid on the response of the trading partners, who are allowed to withdraw "substantially equivalent concessions" of their own. And accordingly, the country initiating the renegotiation (ie, the country desiring less trade volume at the existing terms of trade) can anticipate that it can achieve its desired trade volume at the existing terms of trade. This suggests in turn that incentives to renegotiate a GATT tariff binding will arise any time that, at the existing terms of trade, some government desires less trade volume than its existing tariff commitments imply.

This second application of reciprocity points to a potentially attractive feature of the political optimum: the political optimum is the only point on the efficiency frontier where each government has achieved its preferred local price—and hence its desired trade volume—at the existing terms of trade. At all other points on the efficiency frontier, some government would want more trade volume at the existing terms of trade and some government would want less. In light of the explicit provisions that govern renegotiations of GATT tariff bindings, the political optimum is therefore also the only point on the efficiency frontier where governments would have no reason to attempt to renegotiate their tariff commitments. In this sense, once achieved, the political optimum can be viewed as a particularly robust and stable bargaining outcome of GATT tariff negotiations.

Fig. 2 (adapted from Bagwell and Staiger, 1999) illustrates the point. With the home tariff τ on the vertical axis and the foreign tariff τ^* on the horizontal axis, Fig. 2 depicts the locus of efficient tariff combinations labeled as *EE*. The point on *EE* labeled as *PO* is the political optimum. At the political optimum the iso-welfare contours of the home country (labeled *W*) and the foreign country (labeled *W**) are tangent to each other *and* to the iso-terms-of-trade locus passing through the point *PO* and labeled p^w_{PO}, reflecting the property that each government has achieved its preferred local price and desired trade volume at the existing terms of trade. Now consider the possibility of renegotiation subject to reciprocity beginning from the political optimum. As we have observed, in such renegotiations any country desiring less trade volume at the existing terms of trade can achieve the trade volume it desires; but beginning from the point *PO*, each country already achieves its desired trade volume at the existing terms of trade. Clearly then, and as Fig. 2 illustrates, there is nothing to be gained for either government from such renegotiations: once achieved the political optimum is robust to the possibility of renegotiation subject to reciprocity.

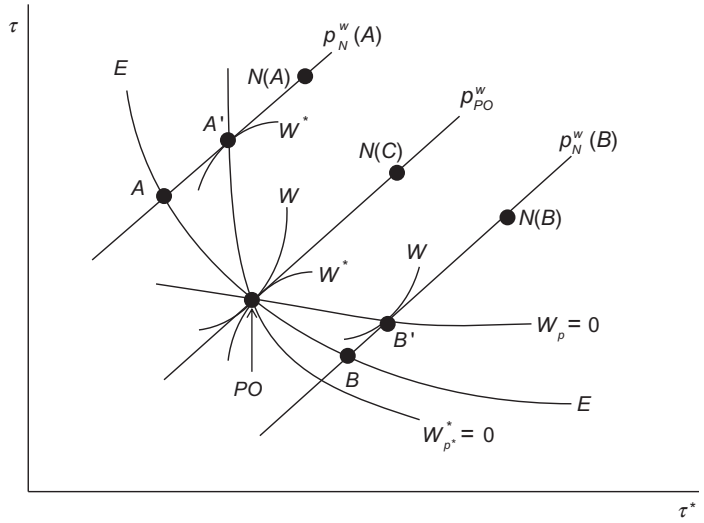

Fig. 2 Renegotiation under reciprocity.

Next consider a point on *EE* that does not correspond to the political optimum, such as point *B* in Fig. 2. Point *B* is also efficient, but relative to the political optimum point *PO* the foreign tariff is higher while the home tariff is lower, with these tariff adjustments orchestrating a terms-of-trade improvement for the foreign country which efficiently transfers surplus toward the foreign country relative to the political optimum. Beginning from an efficient point such as *B*, the possibility of renegotiation subject to reciprocity will be exercised. This is because in such renegotiations, the home government can be assured of achieving no less than its welfare level at the point B' in Fig. 2, where it has achieved its desired trade volume at the existing terms of trade (corresponding to a point on the home government's politically optimal reaction curve where $W_p = 0$); and this is a welfare level that is higher than the welfare level that the home government achieves at the efficient point *B*. Analogous statements apply to an efficient point such as *A* and its counterpart A' in Fig. 2, with the roles of the home and foreign governments reversed. Hence, in light of the explicit provisions that govern renegotiations of GATT tariff bindings, any point on the efficiency frontier *EE* other than the efficient political optimum *PO* is susceptible to renegotiation. And for this reason we can think of the points B', *PO* and A' in Fig. 2 as tracing out the "reciprocity-constrained" efficiency frontier for GATT tariff negotiations.

We have described these two properties of the applications of reciprocity in the GATT/WTO within our benchmark two-good two-country perfectly competitive general equilibrium trade model, but these properties have also been shown to hold in each of the other models described in Section 2 that fall within the terms-of-trade

theory class. Consider for example the monopolistic competition model of firm delocation presented in Section 2.[be] As demonstrated in Bagwell and Staiger (2015), in this model as well these applications of reciprocity both provide a recipe for mutually advantageous liberalization and suggest that, once achieved, the political optimum can be viewed as a particularly robust and stable bargaining outcome of GATT tariff negotiations.

To illustrate these features in the monopolistically competitive model of firm delocation, we follow Bagwell and Staiger (2015) and again define tariff changes that conform to reciprocity as those that bring about equal changes in the volume of each country's imports and exports when valued at existing world prices, but in this partial equilibrium setting we now also take into account trade in the numeraire good. Letting a superscript "0" denote original trade tax levels and a superscript "1" denote new trade tax levels, Bagwell and Staiger show that in this model tariff changes conforming to reciprocity must satisfy[bf]

$$[p^w(\tau_f^0) - p^w(\tau_f^1)]M(p_f^1, p_h^{*1}) = [p^{*w}(\tau_h^{*0}) - p^{*w}(\tau_h^{*1})]E(p_f^1, p_h^{*1}). \tag{73}$$

As with the benchmark competitive model, in the monopolistically competitive model tariff changes that conform to reciprocity remove terms-of-trade manipulation from the calculus of preferred tariffs. Here, according to (73), such tariff changes imply either that world prices are left unchanged as a result of the tariff changes, or if world prices are altered, that they are altered in a way that leaves net trade–tax revenue unchanged.

With the terms-of-trade fixing property of reciprocity now recorded for the monopolistically competitive model, it follows that, starting at the Nash equilibrium, the home and foreign countries must each gain from an adjustment in trade taxes that reduces total trade barriers (t and t^*, and hence by (25) and (27), p_h^* and p_f) and satisfies reciprocity, as long as the reduction in total trade barriers is not too large. Consider for example a small reduction in t and t^* that is brought about by reciprocal reductions in the home and foreign export taxes τ_h^* and τ_f from their Nash levels.[bg] From (73) it follows that the reduction in τ_f that is required to satisfy reciprocity in response to a small reduction in τ_h^*, which we denote by $\frac{d\tau_f}{d\tau_h^*}\Big|_{rec}$, is defined by

[be] Bagwell and Staiger (2015) also demonstrate that these same properties of reciprocity hold in the Cournot model of firm delocation with segmented markets first introduced in Venables (1985). See Bagwell and Staiger (2012b) and (2001b), respectively, for the demonstration that these properties of reciprocity hold in the profit-shifting model of trade agreements and the competitive partial equilibrium model of trade agreements described in Section 2.

[bf] The steps to derive (73) use the balanced trade condition at the original and the new world prices, and are identical to those described in note 19 of Bagwell and Staiger (2001b). See also Bagwell and Staiger (2012b) for a related application.

[bg] See Bagwell and Staiger (2015) for the case of a reduction in t and t^* that is engineered with a small reduction in the home and foreign import tariffs τ_h and τ_f^*.

$$\frac{d\tau_f}{d\tau_h^*}\Big|_{rec} = \frac{E^0}{M^0}, \tag{74}$$

where M^0 and E^0 denote the initial levels of home-country imports and exports, respectively.[bh] But then, evaluated at the Nash conditions given by (36) and (37) and using (74), the impact on home and foreign welfare of a small reciprocal reduction in τ_h^* and τ_f is given, respectively, by

$$-\{V_{p_f}\frac{dp_f}{d\tau_f}\frac{d\tau_f}{d\tau_h^*}\Big|_{rec} + V_{p_h^*}\frac{dp_h^*}{d\tau_h^*} + V_{p^{*w}}\frac{dp^{*w}}{d\tau_h^*} + V_{p^w}\frac{dp^w}{d\tau_f}\frac{d\tau_f}{d\tau_h^*}\Big|_{rec}\} = E^N\frac{dp^w}{d\tau_f} > 0, \text{ and}$$

$$-\{V_{p_h^*}^*\frac{dp_h^*}{d\tau_h^*} + V_{p_f}^*\frac{dp_f}{d\tau_f}\frac{d\tau_f}{d\tau_h^*}\Big|_{rec} + V_{p^{*w}}^*\frac{dp^{*w}}{d\tau_h^*} + V_{p^w}^*\frac{dp^w}{d\tau_f}\frac{d\tau_f}{d\tau_h^*}\Big|_{rec}\} = E^N\frac{dp^{*w}}{d\tau_h^*} > 0.$$

Consider next the impact of reciprocity in the monopolistically competitive model when reciprocity is applied in response to the reintroduction of trade barriers. As with our benchmark competitive model, if countries negotiate to the political optimum, then neither country can gain from unilaterally raising its import tariff or export tax as long as such behavior would result in a reciprocal action from its trading partner, and hence the political optimum is robust to the possibility of renegotiation subject to reciprocity. To see this, suppose we begin at the politically optimal policies defined by (42), and let us focus again on export policies.[bi] Beginning from the political optimum, if the home country were to raise τ_h^* and the foreign government were to reciprocate according to $\frac{d\tau_f}{d\tau_h^*}\Big|_{rec}$, the impact on home-country welfare would be

$$V_{p_f}\frac{dp_f}{d\tau_f}\frac{d\tau_f}{d\tau_h^*}\Big|_{rec} + V_{p_h^*}\frac{dp_h^*}{d\tau_h^*} + V_{p^{*w}}\frac{dp^{*w}}{d\tau_h^*} + V_{p^w}\frac{dp^w}{d\tau_f}\frac{d\tau_f}{d\tau_h^*}\Big|_{rec} = V_{p_f}\frac{dp_f}{d\tau_f}\frac{d\tau_f}{d\tau_h^*}\Big|_{rec} + V_{p_h^*}\frac{dp_h^*}{d\tau_h^*} = 0,$$

where the first equality uses (74) and the fact that $\frac{dp^{*w}}{d\tau_h^*} = \hat{p} = \frac{dp^w}{d\tau_f}$, and the second equality follows according to the conditions for the home-country's politically optimal tariff choices given in (42). An analogous argument applies for the foreign country's incentive to raise τ_f in the face of a reciprocal response from the home country.

Finally, it is straightforward to show that, as in the competitive benchmark model, in the monopolistically competitive model the political optimum is the only point on the efficiency frontier that exhibits this robustness feature. Hence, like the competitive benchmark model, the monopolistic competition model of firm delocation (as well as

[bh] The expression in (74) may be derived in the same way as (73) by considering small tariff changes and dropping second-order terms, and using $\frac{dp^w}{d\tau_f} = \hat{p} = \frac{dp^{*w}}{d\tau_h^*}$.

[bi] See Bagwell and Staiger (2015) for the arguments related to import tariffs.

the Cournot firm delocation model and the profit-shifting model) suggests that the points B', PO and A' in Fig. 2 can be thought of as tracing out the (reciprocity constrained) efficiency frontier for GATT tariff negotiations.[bj]

3.1.3 The Political Optimum as a Focal Outcome of GATT/WTO Negotiations

Our preceding discussion indicates that under GATT Article XXVIII governments will end up on the reciprocity-constrained efficiency frontier as a result of their GATT tariff negotiations. And as Fig. 2 depicts, this implies an efficiency penalty away from the political optimum. The political optimum may therefore be viewed as a possible focal outcome for negotiations of reciprocity-constrained efficient points, in addition to being an idealized outcome for the purpose of undoing the terms-of-trade problem. According to this perspective, deviations from the political optimum to other points on the reciprocity-constrained efficiency frontier are then desirable only to the extent that the implied redistribution of the gains from tariff bargaining across countries makes such deviations desirable, where this redistribution is effected through the implied movements in the terms of trade away from the terms of trade associated with the political optimum.

Is there evidence that the political optimum describes an outcome which is viewed as focal in GATT/WTO discussions, or which is actually delivered in GATT/WTO negotiations, as our theoretical discussion earlier suggests might be the case? This is an important question whose answer has practical implications. It is related to the often-heard claim that the rules-based outcomes of the GATT/WTO protect small and weak countries from the exploitation they would otherwise suffer at the hands of larger and stronger countries if trade negotiations took place outside the GATT/WTO forum, a claim that if true would help to strengthen the link suggested above between GATT's adoption of a common terms of exchange (from which, according to the terms-of-trade theory, the features we have emphasized here then follow) and fairness in the context of trade agreements.[bk] And it is related to the policy debate concerning the performance of the WTO in general and the benefits for developing countries of GATT/WTO membership in particular.

Some broadly suggestive evidence on the focal nature of the political optimum in GATT/WTO discussions can be found in the many references to the GATT/WTO

[bj] See Bagwell and Staiger (2012b, 2015) for further discussion.

[bk] On the rules-based nature of GATT/WTO bargaining, see for example Jackson (1969, pp. 85–86) and Bagwell and Staiger (1999). As we discuss further below, Bagwell and Staiger suggest an added efficiency benefit that, in guiding governments toward the political optimum, GATT's reciprocity norm may help powerful countries commit not to exploit weaker countries at the tariff bargaining table, and in this way encourage the participation of weaker countries in GATT/WTO rounds of negotiation. On whether developing countries benefit from GATT/WTO membership, see for example Jawara and Kwa (2003) and the review of Jawara and Kwa in Staiger (2006), and see Bagwell et al. (forthcoming) for a more general assessment of issues relating to the WTO's performance.

as a "member-driven" organization that leaves room for countries with different needs to pursue different trade policies. Certainly this would be a necessary feature of an institution designed to deliver the political optimum—and thereby to simply strip away the terms-of-trade motivations from the policy choices of governments with diverse policy goals—rather than, say, laissez faire. A suggestive discussion in this regard can be found in the Haberler Report's description of GATT's accommodation of the different needs of industrialized and developing country members:

> *We recognize that there are special considerations affecting the position of under-developed primary producing countries which justify a rather greater use of trade controls by them than by the highly industrialized countries. Industries may need special promotion during the first stages of industrialization when the process of learning industrial techniques is in its early stages and when the promotion of one industry may set a general background which is favourable to the successful growth of others. Special measures to promote industry may be desirable in order to bring into productive employment labour which is under-employed in agriculture. Where the whole or the greater part of a luxury product comes from imports, the restriction of such imports may be administratively the best way of discouraging luxury consumption and promoting savings. Under-developed primary producing countries may be more likely to be in genuine balance-of-payments difficulties than the majority of highly industrialized countries, in which case they will need more often to control imports on these grounds. Finally, insofar as import restrictions can turn the international terms of trade in favour of the restricting countries, it can be argued that poorer countries should have a somewhat greater freedom in their use than richer countries (GATT, 1958, p. 125).*

Rather than advocating laissez faire policies for all, the Panel of Experts for the Haberler Report appears instead to offer as a benchmark a member-driven set of policies where an expansive list of motives for trade controls are seen as potentially justified depending on the needs of individual countries as perceived by their governments, but where terms-of-trade motives are absent from this list.[bl] And with this benchmark established, the

[bl] The excerpt quoted above lists a number of potential rationales for protection that might be interpreted as consistent with the political optimum, but it does not explicitly mention the use of import protection for distributional purposes. This use, too, however, appears to be acknowledged in the Haberler Report, as is confirmed by the following qualification to the Report's statement that a fall in the world price of its import good caused by the policy interventions of its trading partners will benefit an importing country:

> It will, of course, tend to depress the market for the domestic producers in the importing country; but this tendency *can* be offset by protective measures in the importing countries which support the incomes of their producers. The importing country, in any case, purchases its imports on better terms (GATT, 1958, p. 93, footnote 1, emphasis in the original).

On the other hand, it should be noted that while the Report lists a number of points that could legitimize greater use of trade controls in less developed as compared to highly industrialized countries, it also emphasizes that "...there have certainly been cases in which the trade control policies of the under-developed countries have gone far beyond these points in discouraging exports of their primary products and in encouraging import-competing industries" and notes that these policies have probably hurt both the less-developed and the industrialized countries.

potential desirability of a movement away from the benchmark trade policies in the direction that permits terms-of-trade considerations to reenter the determination of policies so as to improve the terms of trade of developing countries and thereby transfer surplus from richer to poorer countries is then suggested for consideration.[bm]

As to whether politically optimal outcomes are actually delivered in GATT/WTO negotiations, at a broad level there are features of the outcomes that have emerged from nearly 70 years of GATT/WTO negotiations that are consistent with the political optimum. For example, it is the policies of the countries which are large in world markets that should be most constrained by a trade agreement that delivers the political optimum, with the policies of the smallest countries left largely or even completely unconstrained. Consistent with this feature, most of the significant market-access concessions have been made by the large industrialized countries, a feature that is in some sense almost guaranteed by the exporter-driven process through which governments identify the markets of their trading partners which they would most like GATT/WTO negotiations to open, and which is further accentuated by special and differential treatment clauses that exempt developing countries and especially the least developed countries from a host of GATT/WTO obligations to which other countries must conform.[bn] And even where a country that is small in a given market accepts on paper obligations that apply to that market, the GATT/WTO enforcement procedures operate "on demand," and so a small player in a market can often expect to violate GATT/WTO obligations in that market with impunity in any event. But beyond these broad observations, there is little available evidence one way or another as to whether GATT/WTO outcomes are well described by the

[bm] Interestingly, in its discussion of agricultural export subsidies the Haberler Report develops this particular point further and advocates that, where possible, the best form for aid from industrialized to less-developed countries is direct financial aid, rather than aid orchestrated indirectly through the terms-of-trade consequences of trade policy intervention and thereby attached to transactions in particular commodities, summarizing with:

> For these reasons we would wish to combine our general recommendation in favour of a moderation of policies for agricultural protectionism with the recommendations (i) that such policies should be moderated in importing and exporting countries simultaneously; and (ii) that economic aid from the richer and more developed to the poorer and less-developed countries—which, as we have argued above (paragraph 148), is certainly much needed—should as far as possible take the form of direct financial aid from the former to the latter. In this way aid can flow in the desired direction and at the same time production can be undertaken where it is most economic. Aid which is indirectly attached to transactions in particular commodities will inevitably make it difficult for the world resources to be used in the most productive manner... (GATT, 1958, pp. 96–97).

From the perspective of Fig. 2, this statement can be interpreted as an argument that a position such as B' can be improved upon by eliminating the use of terms-of-trade movements for purposes of aid to the foreign country, repositioning trade policies at the political optimum point PO, and providing aid directly to the foreign country in the form of direct financial transfers.

[bn] Of course, to the extent that special and differential treatment clauses in the GATT/WTO lead developing countries with significant market power, such as the emerging economies of the BRICS, to not engage in meaningful tariff cutting negotiations of their own in the GATT/WTO, this would interfere with attainment of the political optimum, a point emphasized by Staiger (2006) and Bagwell and Staiger (2014).

political optimum, a fact that is surprising in light of the potential importance of such evidence both in relation to the theory and to the policy debate. Here we discuss several papers that provide some partial evidence on this question.

One way to shed light on this question is to look for evidence that GATT/WTO negotiating outcomes differ significantly from "power-based" bargaining outcomes such as those predicted by the Nash bargaining solution. Baldwin and Clarke (1987) offer an early attempt to gauge the ability of the Nash bargaining solution to track actual bargaining outcomes in the GATT/WTO, focusing on the Tokyo Round of GATT negotiations concluded in 1979. As Baldwin and Clarke note, in the Tokyo Round and the earlier GATT Kennedy Round, an initial bargain occurred over the formula by which each county would in principle cut its tariffs; and then once the formula was agreed, countries engaged in essentially bilateral tariff-item-by-tariff-item negotiations over *exceptions* to the formula cuts. Baldwin and Clarke focus only on the outcomes for the US and the (then) European Community, so their results cannot speak directly to our question here. But their findings are still interesting for the present discussion. They find that the Nash bargaining solution does reasonably well at matching the tariff cuts that would have been implemented under the negotiated formula, but that the Nash bargaining solution performs relatively poorly in predicting the actual tariff cuts that emerged from the Tokyo Round (ie, the cuts that resulted once the exceptions to the formula cuts were taking into account). Baldwin and Clarke conclude (p. 281) that the while concepts such as the Nash bargaining solution "may be successful at predicting the outcome of formal GATT negotiations, they are less accurate in describing the subsequent, more political, process of requesting exceptions from the general formula." Interestingly, Baldwin and Clarke (p. 282) also observe that "[m]any of the pullbacks in both the Kennedy and Tokyo Rounds were on items on which there was not much domestic political pressure for less-than-formula cuts, and hence, were made for reciprocity purposes only... The main point to emphasize is that the process of determining exceptions and achieving reciprocity resulted in a considerably different outcome than would have occurred if the cutting rules were followed strictly." This suggests the possibility that GATT's reciprocity principle may have played an important role in moving the negotiating outcomes of the Tokyo Round away from power-based outcomes such as those predicted by the Nash bargaining solution. What is left unanswered by the Baldwin and Clarke analysis is whether the Tokyo Round outcomes were moved in the direction of the political optimum.[bo]

[bo] Ossa (2014) undertakes a quantitative multicountry analysis of the potential losses from trade wars and potential gains from trade talks while incorporating into his model both inter- and intra-industry trade as well as political economy motives for trade policy intervention. Ossa's analysis of negotiated tariffs focuses on the Nash bargaining solution, but he does not offer a comparison of model predictions under the Nash bargaining solution relative to the actual tariff levels that have emerged from GATT/WTO tariff negotiations (what Ossa terms the "factual" tariffs). Ossa does consider how the MFN principle alters the bargaining outcomes predicted by his model relative to the model's predicted Nash bargaining outcomes in the absence of MFN. We discuss further some of Ossa's findings later in our chapter in the context of our consideration of MFN.

Some suggestive evidence that the political optimum is useful for understanding GATT/WTO negotiating outcomes is provided from a different perspective by the findings of Bagwell and Staiger (2011) and Bagwell et al. (2015). Bagwell and Staiger study the tariff cuts agreed in WTO accession negotiations by 16 countries that joined the WTO as new members after its creation in 1995. They ask whether the pattern of agreed tariff cuts for these countries correspond to the pattern that would be predicted by the terms-of-trade theory if the cuts moved each acceding country from its reaction curve (pre-WTO membership) tariff to its politically optimal reaction curve, and find strong support for this prediction. Bagwell et al. study the recently declassified GATT bargaining records from 1950 to 1951 Torquay Round. They ask whether the pattern of tariff cuts offered by the US in this round, and the probability that the US offers were successful (ie, led to an agreed US tariff binding in the round), can be understood from the perspective of the terms-of-trade theory under the assumption that the US offers correspond to its politically optimal reaction curve tariffs.[bp] Bagwell et al. too, find support for the terms-of-trade theory and the view that the US made tariff offers at Torquay that resided on its politically optimal reaction curves. These two papers do not yield direct evidence on whether GATT/WTO outcomes deliver the political optimum, because they merely suggest that the tariff bindings of individual countries agreed in GATT/WTO negotiations can be understood with reference to their politically optimal reaction curves, but the findings are nonetheless suggestive. In terms of Fig. 2, the distinction here is between, on the one hand, whether GATT delivers the political optimum as defined by the point *PO*, and on the other hand whether the politically optimal reaction curve as depicted for example by the point B' is helpful for predicting the negotiated tariff choices of individual countries. Our question posed above concerns the first point, but the supporting evidence on the second point provided by these two papers is still encouraging, in the sense that a lack of support on this second point would have cast doubt on an affirmative answer to the question of interest.

3.1.4 Reciprocity and the GATT/WTO Bargaining Process

Thus far we have considered separately the two applications of reciprocity that arise in GATT practice, one relating to market access liberalizing tariff negotiations and the other relating to renegotiation of tariff commitments for the purpose of reintroducing tariffs. The second of these applications of reciprocity constitutes a fairly rigid rule in GATT/WTO practice, but the first is in principle simply a norm of behavior.

There is evidence, however, that supports an important role for even this first application of reciprocity in the GATT/WTO practice. Finger (1979) describes the central role that the desire to achieve reciprocity as felt by GATT member governments played in determining the outcomes of each GATT round through the 1964–67 Kennedy

[bp] As Bagwell et al. (2015) demonstrate and as we describe further below, according to the terms-of-trade theory a country's tariff-cutting offers should correspond to its politically optimal reaction curve under strict adherence to reciprocity and MFN.

Round, focusing on how this desire and the exception from reciprocity granted to less-developed countries prevented the less-developed countries from achieving meaningful increases in access to the markets of their trading partners as a result of these rounds (see also Hoda, 2001, pp. 52–63). And in an early analysis of the 1950–51 Torquay Round prepared for the US International Chamber of Commerce (1955, p. 33), the need for reciprocity in GATT tariff negotiations was viewed as sufficiently rigid to constitute an impediment to further negotiated tariff liberalization in light of existing tariff asymmetries across the bargaining countries. More recently, Shirono (2004) finds that the tariff-cutting results of the Uruguay Round conform well with the reciprocity norm. Focusing on US tariff cuts in the Uruguay Round and constructing a measure of market-access concessions while instrumenting to address the potential endogeneity issues, Limão (2006, 2007) also finds evidence consistent with reciprocity, reporting that a decrease in the tariff of a US trading partner that exports a given product leads to a decrease in the US tariff on that product, and that a significant determinant of cross-product variation in US tariff liberalization is the degree to which the US received reciprocal market-access concessions from the corresponding exporting countries. A similar exercise for the EU tariff cutting behavior in the Uruguay Round is carried out by Karacaovali and Limão (2008), who find analogous support for the importance of reciprocity in explaining the pattern of EU tariff cuts. Evidence of reciprocity may be stronger for some sectors than others, however. Examining tariff liberalization by the US in the Uruguay Round, Gulotty (2014) reports evidence that sectors with high contract-intensive products do not exhibit reciprocity.

While more evidence is needed and findings may differ across sectors, the empirical work described above is broadly supportive of an important role for the reciprocity norm in actual GATT/WTO tariff negotiations. In combination with our earlier observation that the application of reciprocity in the context of renegotiations constitutes a fairly rigid rule in GATT/WTO practice, it is then relevant to consider the ramifications for the GATT/WTO bargaining process of a rigid application of reciprocity *both* in the context of tariff liberalizing negotiations *and* in the context of renegotiation to higher tariff levels. To this end, we return to Fig. 2, and illustrate a simple point: the first application of reciprocity focuses bargaining on the volume of trade rather than the terms of trade; and together with the first application, the second application of reciprocity in GATT effectively eliminates (strategic) bargaining on the volume of trade as well. In fact, as we now describe, together the two applications of reciprocity in GATT simplify the market access bargaining process and leave each country with a dominant "truth-telling" strategy to offer to adopt its politically optimal reaction curve tariffs in exchange for reciprocal cuts from its trading partners.[bq]

[bq] For a more complete presentation of these points, see Bagwell and Staiger (1999) and Bagwell et al. (2015). Bagwell et al. extend these arguments to a multicountry setting in the presence of MFN; we discuss their extended results further in the next section.

That the first application of reciprocity focuses bargaining on the volume of trade rather than the terms of trade follows directly from the terms-of-trade fixing property of reciprocity highlighted above. In terms of Fig. 2, this is manifested in the fact that, beginning from any pair of initial tariffs, tariff liberalizing negotiations that conform to reciprocity must lead to an outcome on the iso-terms-of-trade locus passing through the initial tariff pair. In Fig. 2 we take the initial tariffs to be the pair of Nash tariffs and illustrate with three possible Nash points labeled in the figure as $N(A)$, $N(B)$ and $N(C)$. Notice that, beginning from any of these Nash points and with the terms-of-trade pinned down by the requirement of reciprocity, the preferred trade volume of each government corresponds to the point on the government's politically optimal reaction curve at this terms-of-trade (ie, the tangency of the government's iso-welfare contour with the relevant iso-terms-of-trade locus). Any bargaining that occurs in the presence of this first application of reciprocity would then be over trade volumes at the fixed terms of trade, with each government attempting to achieve its preferred volume. Next notice from Fig. 2 that, together with the first application, the second application of reciprocity in GATT effectively eliminates (strategic) bargaining on the volume of trade as well. This is because under the second application of reciprocity, no government can be forced to accept more trade volume than it desires at the existing terms of trade, and so the government whose preferred volume is lowest will ultimately get its way.

As Bagwell and Staiger (1999) argue, the implications of the two applications of reciprocity in GATT can together then be captured in the following stylized game. An initial pair of tariffs, say the Nash point $N(C)$ in Fig. 2, corresponds to a particular iso-terms-of-trade line. Governments simultaneously make tariff proposals, where any proposal must conform to reciprocity and therefore amount to a tariff pair that lies along the iso-terms-of-trade line passing through the point $N(C)$. If the proposals agree, then the common proposal is implemented, while if the proposals disagree then the proposal implying the least trade volume (the higher tariff pair) is implemented. In this game, as Bagwell and Staiger demonstrate, it is a dominant strategy for each government to propose the tariff pair that if implemented would deliver its preferred trade volume at the existing terms of trade (ie, its politically optimal reaction curve tariff paired with the reciprocity-consistent tariff for its trading partner). And importantly, it is easy to see that this conclusion holds whether or not governments possess private information about their preferred trade volumes.

This discussion suggests a potential benefit from strict adherence to reciprocity in the two applications where it arises in GATT: strict adherence to reciprocity can induce governments to reveal their politically optimal reaction curves, and thereby allow governments to avoid costly bargaining delays that might otherwise be associated with the presence of private information (see, for example, Admati and Perry, 1987 and Cramton, 1992). But the costs of strict adherence to reciprocity are also evident from Fig. 2: whenever the negotiating environment is asymmetric, in the particular sense that

the Nash terms of trade differ from the politically optimal terms of trade, negotiations under strict adherence to reciprocity fails to reach the efficiency frontier. In Fig. 2 this is illustrated by the fact that the stylized game described just above would deliver governments to the political optimum in the symmetric case corresponding to the Nash pair $N(C)$, where the Nash and the politically optimal tariffs both lie on the same iso-terms-of-trade line.[br] Asymmetric cases are illustrated in Fig. 2 by the Nash points $N(A)$ and $N(B)$:[bs] for these asymmetric cases, the stylized game described earlier would deliver governments to the inefficient points A' and B', respectively, in Fig. 2. In general, as Fig. 2 illustrates, the performance of strict adherence to reciprocity in GATT/WTO tariff negotiations as measured by its ability to deliver governments to the efficiency frontier deteriorates with the asymmetry of the environment.

As is known from the Myerson and Satterthwaite Theorem (Myerson and Satterthwaite, 1983), in the presence of private information it is generally not possible to design an institution or mechanism that satisfies certain attractive constraints (Bayes–Nash incentive compatibility, interim individually rationality and budget balancing) and yet achieves ex-post efficient (first-best) outcomes. Still, given these constraints, the optimal or second-best mechanism would typically out-perform the strict adherence to reciprocity that we have described earlier. But in this regard the observations of Hagerty and Rogerson (1987) concerning shortcomings of this Bayesian approach to second-best mechanisms seem relevant:

> *The key shortcoming of this approach is that it relies heavily on the assumption that there exists a common prior over traders' valuations known to all participants. In particular, an institution which produces a very efficient outcome for one prior might perform very poorly under some other prior. This creates two related problems. First, a social planner may not be able to ascertain exactly what traders' priors are when choosing an institution. Second, given the costs of creating new institutions, a trading institution (such as a stock exchange, for example) is often chosen with the intention that it will be used by a variety of traders over a long period of time. A variety of priors might be expected to occur over this time. These problems suggest that an important concern when choosing a trading institution is that it work "fairly well" over a broad range of priors, ie, that it be robust with respect to changes in the information structure of the market (Hagerty and Rogerson, 1987, p. 95).*

On this basis, Hagerty and Rogerson advocate the consideration of mechanisms under which each player has a dominant strategy, which they argue can then avoid these

[br] In the symmetric case, beginning from their politically optimal tariffs at the point PO in Fig. 2, if a trade war were to break out and move the countries back to the Nash point $N(C)$, neither country would succeed in pushing the terms-of-trade in its favor.

[bs] In these asymmetric cases, if a trade war were to break out beginning from PO the home country would succeed in pushing the terms of trade in its favor in the case corresponding to the Nash point $N(A)$ and the foreign country would succeed in pushing the terms of trade in its favor in the case corresponding to the Nash point $N(B)$.

shortcomings. Interestingly, Hagerty and Rogerson show for the simple bilateral trade setting that "posted-price" mechanisms, whereby a price is posted in advance and trade occurs if and only if all traders agree to trade, are essentially the only mechanisms such that each trader has a dominant strategy. The general-equilibrium model that we consider differs from a simple bilateral trade setting. Nevertheless, together the two applications of reciprocity in GATT can be interpreted as working like a posted price mechanism for tariff bargaining and therefore exhibiting these robustness benefits.[bt] From this perspective, strict adherence to reciprocity may be understood to be a potentially attractive design feature of the GATT/WTO.

It thus appears that, in the context of trade agreements and from the perspective of the terms-of-trade theory, fixing the terms of exchange of market access at a common level for all countries—from which GATT's one-for-one definition of reciprocity and each of the features we have described earlier then follows—produces a number of potentially desirable properties, and the more so the more symmetric is the underlying tariff bargaining environment. Notice, too, that these properties should remain desirable as long as terms-of-trade manipulation is an important problem for trade agreements to solve, even if it is not the *only* problem as the terms-of-trade theory suggests.

We close this section with a brief consideration of a question we have thus far ignored: Why are the terms of exchange of market access in the GATT/WTO expressed in barter terms; that is, why is there no role for cash? This is an important question that has received little attention in the trade agreements literature. Greater use of cash transfers in the WTO system offers potential benefits. As the Haberler Report well describes (see note bm), cash payments may be attractive as a relatively efficient instrument with which to achieve transfers across countries.[bu] From this perspective, monetary compensation is a potentially attractive means of offering compensation in bilateral disputes.[bv] Furthermore, and especially as asymmetries in the GATT/WTO system become increasingly prominent (for example, between the advanced industrialized countries and the BRICS), the findings we have described earlier suggest that the GATT/WTO reliance on reciprocity could become increasingly problematic, and other methods of orchestrating

[bt] An interesting direction for future research concerns the conditions under which dominant strategy implementation in the general-equilibrium model of trade necessitates that an exchange of tariff changes must satisfy the principle of reciprocity.

[bu] As Bagwell and Staiger (2005a) confirm, in a partial-equilibrium model with privately informed governments, governments can achieve first-best efficient policies when cash transfers are available.

[bv] In a self-enforcing, repeated-game context, Limão and Saggi (2008) argue that a system in which retaliatory tariffs enforce the payment of monetary fines does not offer greater cooperation than a system that relies directly on retaliatory tariffs. A system with fines offers an advantage when (unanticipated) shocks result in actual disputes, however. Bagwell et al. (2007) demonstrate that the auctioning of retaliation rights in the GATT/WTO could serve as a third-party mechanism for enforcing the payment of cash compensation to countries injured by GATT/WTO inconsistent policies of their trading partners.

internationally efficient trade liberalization—such as holding auctions for reductions in tariffs and increased market access in exchange for cash payments—may in principle become relatively more attractive.

At the same time, the use of cash transfers raises potential concerns as well. One potential drawback is suggested by the finding in Bagwell and Staiger (1999) that, in guiding governments toward the political optimum and away from power-based outcomes, GATT's reciprocity norm may help more powerful countries commit not to exploit weaker countries at the tariff bargaining table, and thereby help to solve a potential hold-up problem that could otherwise reduce the participation of weaker countries in GATT/WTO rounds of negotiation. The availability of monetary transfers at the negotiation stage would make such commitments less meaningful, because direct monetary transfers can be used to undo the transfers implied by the terms-of-trade movements of tariff choices; and this could be one potential downside of facilitating a greater role for such transfers in exchange for trade liberalization. In the broader bargaining literature, Harstad (2007) makes the argument that the availability of cash side-payments in settings of bargaining with private information can exacerbate the reasons for delay and lead to worse outcomes in some situations than if the cash payments were not available. We discuss further in the context of our consideration of MFN an additional potential reason suggested by the literature that cash payments are not more prominent in the GATT/WTO, but we see this as an understudied and fruitful area for further research.

3.2 Reciprocity and the Terms-of-Trade Theory Under Alternative Assumptions

We now return to the first application of reciprocity in the GATT/WTO with which we began our discussion, namely, the application that arises when negotiations are aimed at liberalizing market access as in multilateral GATT/WTO negotiating rounds under GATT Article XXVIII bis. We described in our earlier discussion how the structure of international cost-shifting implies that, beginning from their Nash tariff choices, each government would desire greater trade volume if this greater trade volume could be achieved at a fixed terms of trade. And we argued that tariff-cutting according to reciprocity is a way to achieve such fixed-terms-of-trade increases in trade volume, and that liberalization under the principle of reciprocity thus delivers mutual gains.

Here we comment on how this conclusion must be modified when a number of the standard assumptions that usually accompany the terms-of-trade theory of trade agreements are relaxed. And for this purpose we now return to our benchmark two-good two-country perfectly competitive general equilibrium trade model described in Section 2.

A first and critical assumption is contained in (5), which states that $W_{\tilde{p}^w} < 0$ and $W^*_{\tilde{p}^w} > 0$ and indicates that each government benefits from a terms-of-trade improvement, when the local price in its country is held fixed. As we observed, this assumption is satisfied by each of the leading models of trade policy formulation. But as Blanchard

(2007a, 2010) shows, there is an important qualification to this statement that can arise in the presence of international ownership. Blanchard (2007a) considers the impact that export-platform foreign direct investment can have on the unilateral tariff choices of the investment-source country, while Blanchard (2010) extends the analysis of optimal tariffs to international ownership more generally. In both cases, Blanchard demonstrates that international ownership can mitigate the unilateral incentive to use tariffs for purposes of manipulating the terms of trade, and thereby can have important effects on the inefficiency of unilateral tariff setting, serving to reduce Nash tariffs closer to their internationally efficient levels and possibly even to lead to Nash tariffs that are efficient (in a knife edge case) or too *low* from an international perspective. In relation to the Nash conditions described by (6), the key difference here is that international ownership operates to diminish the absolute value of $W_{\widetilde{p}^w}$ and $W^*_{\widetilde{p}^w}$, possibly all the way to zero, and may even reverse the sign of these terms.[bw]

As Blanchard (2007a, 2010) demonstrates and as the Nash conditions in (6) confirm, if international ownership ends up reversing the signs of $W_{\widetilde{p}^w}$ and $W^*_{\widetilde{p}^w}$, then in the Nash equilibrium we would have $W_p > 0$ and $W^*_{p^*} < 0$ and a trade agreement would need to *raise* tariffs to reach the international efficiency frontier; this can be seen most clearly by considering a trade agreement that satisfies reciprocity and therefore fixes $\widetilde{p}^w$, and where it is then clear that both countries gain by reciprocally raising their tariffs. And if the

[bw] In addition to this international-ownership effect, Blanchard (2010) identifies a second effect, according to which the domestic local price can be manipulated to extract rents from foreign investors; and with this second effect she demonstrates that in the presence of international ownership the terms of trade is not the only channel through which externalities travel. As we describe earlier, local price externalities also arise in the profit shifting and delocation models analyzed by Bagwell and Staiger (2012a,b, 2015), but Bagwell and Staiger show for these models that the terms-of-trade externality is nevertheless the only source of policy inefficiency in the Nash equilibrium (ie, at the political optimum these local price externalities are "shut down"). In her working paper (see Blanchard, 2007b, appendix 5.3), Blanchard establishes a related result in the presence of international ownership. Exploiting the fact that market clearing allows p^* to be expressed as $p^*(p,\widetilde{p}^w)$ and p to be expressed as $p(p^*,\widetilde{p}^w)$, thereby allowing the home and foreign objectives to be written as $W(p,\widetilde{p}^w) \equiv w(p,p^*(p,\widetilde{p}^w),\widetilde{p}^w)$ and $W^*(p^*,\widetilde{p}^w) \equiv w^*(p^*,p(p^*,\widetilde{p}^w),\widetilde{p}^w)$, respectively, even in the presence of local price externalities, and defining the political optimum with respect to the W and W^* functions (ie, politically optimal home and foreign tariff selections are made "as if" $W_{\widetilde{p}^w} \equiv 0$ and $W^*_{\widetilde{p}^w} \equiv 0$ respectively), Blanchard shows that at the political optimum the local price externalities vanish—and therefore that the political optimum remains efficient—and in this sense that the terms-of-trade externality remains the only source of policy inefficiency in the Nash equilibrium in the presence of international ownership. It is important to note, however, that the political optimum defined by Blanchard is related to but different than the political optimum as we have defined it above. According to the definition employed by Blanchard, not valuing movements in $\widetilde{p}^w$ implies not valuing the pure international transfers associated with such movements *and* the trade volume changes associated with the implied local price movement (ie, the implied movement of p^* in the case of $W_{\widetilde{p}^w}$ and of p in the case of $W^*_{\widetilde{p}^w}$), whereas according to the definition of the political optimum that we have described earlier it is only the pure international transfer associated with movements in $\widetilde{p}^w$ that is assumed not to be valued.

pattern of international ownership happened to deliver $W_{\tilde{p}^w} = 0 = W^*_{\tilde{p}^w}$ then governments would adopt the efficient politically optimal tariffs (defined by $W_p = 0 = W^*_{p^*}$) in the Nash equilibrium and the pattern of international ownership would obviate the need for a trade agreement completely. More generally, even absent these extreme cases, it is clear that by reducing the absolute magnitude of $W_{\tilde{p}^w}$ and $W^*_{\tilde{p}^w}$, international ownership can reduce the magnitude of the terms-of-trade driven policy inefficiencies associated with unilateral tariff choices.[bx] Finally, whether the pattern of international ownership requires governments to agree to lower or rather to raise their tariffs beginning from Nash to reach the international efficiency frontier, Blanchard argues that the principle of reciprocity continues to serve as an important guide to efficient outcomes, once the definition of market access reflects ownership positions.

A second important assumption that is usually made in the context of the terms-of-trade theory is contained in (4), which states that $\frac{dp}{d\tau} > 0 > \frac{\partial \tilde{p}^w}{\partial \tau}$ and $\frac{dp^*}{d\tau^*} < 0 < \frac{\partial \tilde{p}^w}{\partial \tau^*}$ and indicates that prices respond to tariffs in the "regular" way and therefore do not exhibit either the Lerner or the Metzler paradox. How must the conclusion above be qualified if this assumption fails? An initial observation is that, as long as international markets are integrated, the Lerner and Metzler paradoxes cannot both hold at once.[by] To see this, consider the home country, where existence of the Metzler and Lerner paradoxes would imply $\frac{dp}{d\tau} < 0 < \frac{\partial \tilde{p}^w}{d\tau}$. As the assumption of integrated markets implies $p = \tau \tilde{p}^w$ for nonprohibitive tariffs, it follows immediately that we can not simultaneously have Metzler and Lerner paradoxes in this setting. Hence, there are two remaining cases to consider.

A first case is where the Metzler paradox arises but there is no Lerner paradox. Focusing on the domestic country, this corresponds to the case where $\frac{dp}{d\tau} < 0$ and $0 > \frac{\partial \tilde{p}^w}{d\tau}$. This can happen if a tariff hike pushes $\tilde{p}^w$ down to such an extent as to overwhelm the direct effect on p of the tariff hike. For this case, the domestic government's Nash first-order condition recorded in (6) then implies $W_p > 0$, where we assume that $W_{\tilde{p}^w} < 0$. This means that, under the principle of reciprocity (ie, holding the world price fixed), the domestic government would want to raise $p = \tau \tilde{p}^w$, and similarly for the foreign government.

[bx] Assessing the impact of US multinational firms' affiliate offshoring behavior on US tariff preferences, Blanchard and Matschke (2015) find evidence consistent with this effect. Gulotty (2014), however, argues that greater international ownership does not similarly lead to reductions in regulatory barriers. He argues that regulatory barriers raise fixed costs, and that the associated reduction in entry may lead to net gains for efficient, globalized firms.

[by] With segmented markets, the Lerner and Metzler paradoxes can coexist (see for example the discussion of export policies in the Cournot delocation model in Bagwell and Staiger, 2015).

Therefore, when the Metzler paradox is present, mutual gains under the principle of reciprocity are achieved via reciprocal tariff *increases*. To understand why, consider the externality on W of a change in τ^*:

$$\frac{dW}{d\tau^*} = W_p\frac{dp}{d\tau^*} + W_{\tilde{p}^w}\frac{\partial\tilde{p}^w}{d\tau^*} = [W_p\tau + W_{\tilde{p}^w}]\frac{\partial\tilde{p}^w}{d\tau^*}$$
$$= [\tau(\frac{-W_{\tilde{p}^w}\frac{\partial\tilde{p}^w}{d\tau}}{\frac{dp}{d\tau}}) + W_{\tilde{p}^w}]\frac{\partial\tilde{p}^w}{d\tau^*} = [\tilde{p}^w]W_{\tilde{p}^w}\frac{\partial\tilde{p}^w}{d\tau^*}\frac{1}{\frac{dp}{d\tau}} > 0,$$

where the third equality uses the domestic government's Nash first-order condition and the inequality follows since $\tilde{p}^w > 0$, $W_{\tilde{p}^w} < 0$, $\frac{\partial\tilde{p}^w}{d\tau^*} > 0$ (no Lerner paradox for foreign, either) and $\frac{dp}{d\tau} < 0$. With this, an envelope argument suggests that mutual gains can be achieved starting at Nash if each country slightly increases its tariff, since its trading partner thereby enjoys a first-order positive externality despite the fact that the trading partner suffers a terms-of-trade loss. Intuitively, if home were to cut τ slightly from the Nash level, it would receive a small increase in p, and under the Nash first-order condition the value of this increase would be exactly offset by home's consequent terms-of-trade loss. Let us then fix this p increase and consider an alternative way of achieving it, namely through a small increase in τ^*, which induces a higher p by raising $\tilde{p}^w$. While the same p is achieved under both approaches, notice that $\tilde{p}^w$ is higher under the first approach, since τ is reduced in this scenario (and $p = \tau\tilde{p}^w$ always holds). So, if home were indifferent about achieving a higher p while suffering a high value for $\tilde{p}^w$, then home will strictly benefit by getting the same higher p in the company of a lower value for $\tilde{p}^w$. The externality is therefore positive, even though it generates a terms-of-trade loss for the recipient.[bz]

The second case is where the Lerner paradox arises but there is no Metzler paradox. Focusing again on the domestic country, this corresponds to the case where $\frac{dp}{d\tau} > 0$ and $0 < \frac{\partial\tilde{p}^w}{d\tau}$. For this case, the domestic government's Nash first-order condition in (6) then implies $W_p > 0$, where again we assume that $W_{\tilde{p}^w} < 0$. This means that, under the principle of reciprocity (ie, holding the world price fixed), the domestic government would again want to alter tariffs so as to raise $p = \tau\tilde{p}^w$, and similarly for the foreign government.

For the case where the Lerner paradox is present, therefore, mutual gains under the principle of reciprocity are thus again achieved via reciprocal tariff *increases*. To

[bz] This intuition is similar to that described for the case of export taxes in the linear Cournot delocation model analyzed by Bagwell and Staiger (2012a).

understand why, consider again the externality on W of a change in τ^* beginning from Nash. Following the same steps as above, we find that:

$$\frac{dW}{d\tau^*} = [\widetilde{p}^w] W_{\widetilde{p}^w} \frac{\partial \widetilde{p}^w}{d\tau^*} \frac{1}{\frac{dp}{d\tau}} > 0,$$

where the inequality now obtains since $\widetilde{p}^w > 0$, $W_{\widetilde{p}^w} < 0$, $\frac{\partial \widetilde{p}^w}{d\tau^*} < 0$ (Lerner paradox for foreign, too) and $\frac{dp}{d\tau} > 0$. So, once again, starting at Nash, a higher foreign tariff generates a positive externality for home. In this case, though, the higher foreign tariff also gives home a terms-of-trade gain, due to the Lerner paradox.

It therefore appears that when either the Metzler or the Lerner paradox is present, the terms-of-trade theory's implication that liberalization under the principle of reciprocity delivers mutual gains must be modified: it is still the case that reciprocal tariff changes deliver mutual gains, but now the sign of these changes is reversed, with reciprocal tariff increases pointing the way to the international efficiency frontier. Finally, we note that when the assumption of integrated markets is relaxed and segmented markets are instead considered, the standard reciprocal import-tariff-liberalizing predictions of the terms-of-trade theory can survive both the Metzler and the Lerner paradox, though new predictions can also arise regarding export policies in such settings.[ca]

3.3 Reciprocity Beyond the Terms-of-Trade Theory

In the preceding sections we have considered reciprocity from the perspective of a variety of models that fall within the terms-of-trade theory of trade agreements, as well as a number of extensions to those models. We close this section with a brief discussion of reciprocity from modeling perspectives that fall outside the terms-of-trade theory, including environments where governments have a limited set of trade policy instruments, where offshoring is prevalent, and where governments seek to make commitments to their own private sectors through trade agreements.

A first observation is that, in each of these environments, we may continue to expect that tariff changes conforming to reciprocity as we have defined reciprocity above will hold the terms of trade fixed. This is because, as we noted above, we have derived this property of reciprocity in a modeling framework that features little more than market clearing and trade balance, and that is thus general enough to include each of the environments that we consider below.

[ca] See Bagwell and Staiger (2012a, 2015) and Bagwell and Lee (2015) for theoretical explorations of the design of trade agreements in settings that feature Metzler and Lerner paradoxes, and see Ludema and Zhi (2015) for recent empirical evidence relating to the existence of the Metzler paradox.

We therefore turn to the remaining two questions around which we have organized our discussion of reciprocity above. First, beginning from their Nash tariffs, can both countries gain from reciprocal liberalization in these alternative modeling environments provided they do not go too far? And second, in these environments can the political optimum still be singled out as the only point on the efficiency frontier that is robust to renegotiation subject to reciprocity?

Consider first the case of missing trade policy instruments. For this case we return to the competitive partial equilibrium model with missing instruments described in Section 2, where we assumed that export policies are prohibited so that the home government has only an import tariff τ_x and the foreign government has only an import tariff τ_y^*. In this environment, the Nash tariffs are defined by the first-order conditions

$$\text{Home}: \; W_{\hat{p}_x}\frac{d\hat{p}_x}{d\tau_x} + W_{p_x^w}\frac{\partial p_x^w}{\partial \tau_x} = 0$$

$$\text{Foreign}: \; W^*_{\hat{p}^*_y}\frac{d\hat{p}^*_y}{d\tau^*_y} + W^*_{p_y^w}\frac{\partial p_y^w}{\partial \tau^*_y} = 0.$$

With $W_{\hat{p}_x} < 0$ and $W^*_{\hat{p}^*_y} < 0$ implied, an immediate observation is then that, beginning from their Nash import tariffs, both countries can gain from at least a small amount of reciprocal liberalization, because such tariff reductions lower $\hat{p}_x$ and $\hat{p}^*_y$ without inducing any welfare-relevant changes in p_x^w and p_y^w and therein raise home and foreign welfare according to $W_{\hat{p}_x} < 0$ and $W^*_{\hat{p}^*_y} < 0$. So our earlier answer to the first question above is unchanged by the limitation we have imposed here on the set of trade instruments. But our earlier answer to the second question is overturned in this environment. This follows directly from the result reported earlier, that with missing trade policy instruments it is no longer generally the case that politically optimal policies are internationally efficient, because there is a local price externality that now persists at the political optimum. In fact, as we argued above, when export policies are missing the political optimum will be inefficient whenever export industries enjoy political support in the objective functions of their governments.

We may therefore conclude that in an environment with missing trade policy instruments, the political optimum cannot in general be singled out as the only point on the efficiency frontier that is robust to renegotiation subject to reciprocity, because the political optimum does not itself generally rest on the efficiency frontier in this environment. Moreover, in light of the fact that any efficient point that does not correspond to the political optimum is susceptible to such renegotiation, we can make a stronger statement: in general for this missing-instrument setting, *no* point on the efficiency frontier is robust to renegotiation subject to reciprocity. This suggests that the attractive features of reciprocity that hold in the context of the terms-of-trade theory and that we have emphasized above are substantially diminished in environments where trade policy instruments are missing.

A key question, then, is how to interpret the possibility of missing trade policy instruments. One possible interpretation is that countries do not in fact possess complete sets of trade policy instruments, with a possible candidate for missing instruments being export subsidies whose funding needs might make these policies simply inaccessible to all but the richest countries. Under this interpretation a trade agreement must be designed to both internalize the terms-of-trade externality and provide the missing trade policy instruments. A second possible interpretation is that these trade policy instruments are not truly missing, but rather their use has been prohibited by international agreement. Under this second interpretation the problem to be solved by a trade agreement is still *fundamentally* the terms-of-trade problem, but as part of the approach to solving the terms-of-trade problem the agreement has altered the nature of the policy externalities with which the agreement itself must contend. From this perspective an important question is whether the increasingly stringent prohibition on the use of export subsidies as this prohibition has evolved from GATT to the WTO (see, for example, Sykes, 2005) creates such a missing-instruments environment for WTO member governments; and if so, whether the GATT/WTO traditional reliance on reciprocity is growing increasingly at odds with an institution well designed to solve the terms-of-trade driven Prisoners' Dilemma problem.[cb]

Ossa (2011) provides an interesting answer to this question. He demonstrates that, in a general equilibrium monopolistic competition model of firm delocation where export policies are ruled out, an adaptation of the definition of reciprocity can be found which largely preserves the properties of reciprocity we have emphasized above. Ossa's model has both a manufacturing and a nonmanufacturing sector, with the manufacturing sector composed of many firms producing differentiated manufacturing goods and representing the sector where firm delocation effects occur. In Ossa's model the terms-of-trade manipulation problem is completely absent because, as in the partial equilibrium monopolistic competition model of firm delocation we described in Section 2, only export policies can have terms-of-trade impacts, and as we noted in Ossa's model the use of export policies is ruled out by assumption. This allows Ossa to highlight the firm-delocation problem that is then transmitted across countries though local-price channels. Ossa shows in this setting that if reciprocity is defined to cover only changes in manufacturing exports and imports, and not also changes in the trade of the nonmanufacturing good as would be consistent with our definition of reciprocity above, then tariff changes that conform to this adaptation of reciprocity keep the numbers of firms in each country unchanged and hence will be free of the firm delocation externality that countries would otherwise

[cb] There is also the related question of how to interpret the increasing stringency of GATT/WTO rules on export subsidies and subsidies more generally. On attempts to interpret these developments, see for example Bagwell and Staiger (2001c, 2006, 2012a), Potipiti (2012), Brou and Ruta (2013), and Bagwell and Lee (2015).

impose on one another with their trade policy choices; and for this reason reciprocity in Ossa's model behaves much like the terms-of-trade fixing property of reciprocity that we have emphasized above in the context of the terms-of-trade theory.

In particular, Ossa (2011) demonstrates that, beginning from Nash and under the restriction of nonnegative import tariffs, countries can gain from reciprocal tariff liberalization if they abide by this definition of reciprocity. And he shows that any point on the efficiency frontier is robust to renegotiation subject to reciprocity when reciprocity is defined in this way. More broadly, Ossa's findings point to the possibility that the principle of reciprocity can constitute a sensible design feature of trade agreements in environments that extend beyond the terms-of-trade theory, and at the same time suggest that in the absence of a full set of trade policy instruments it may be important to adopt a flexible view about the precise definition of reciprocity.[cc]

We consider next how the answers to the two questions above must be modified in environments where offshoring is prevalent. To begin, starting from their Nash tariffs, can both countries gain from at least a small amount of reciprocal liberalization? As it turns out, in the presence of offshoring this is no longer assured.

To see why, we return to the model of Antràs and Staiger (2012a). Recall that in (53) we defined the international input price p_x^*; and as the world price of the final good is fixed on world markets by assumption, p_x^* represents the terms of trade between the home and foreign country in this model. Hence, reciprocal reductions in τ_1^H, τ_x^H and τ_x^F are defined by any set of tariff reductions that hold p_x^* fixed. Next recall that Nash policies are indeed inefficient in this model. But as Antràs and Staiger show, the Nash inefficiencies take a particular and interesting form: both the volume of input trade ($\hat{x}^N$) *and* the local price of the final good in the home-country market (p_1^{HN}) are inefficiently low. From this vantage point, it can now be seen that, beginning from the Nash point, even small reciprocal tariff cuts can lead to losses in this setting for a simple reason: if the home

[cc] In this light it is interesting to observe that, while reciprocity in the GATT/WTO has a specific definition in the context of the reintroduction of protective measures and what constitutes a reciprocal response to that reintroduction (see note as), the precise definition of what constitutes reciprocity in the context of negotiations aimed at liberalizing market access, as is the focus of multilateral GATT/WTO negotiating rounds under GATT Article XXVIII bis, has been left up to each country to decide. Hoda (2001) provides an illuminating account:

As mentioned earlier, neither the provisions of GATT 1994 nor the procedures of the eight rounds of tariff negotiations indicate how reciprocity is measured or defined. At the [1955 GATT] Review Session, Brazil had proposed a formula for measurement of concessions for determining reciprocity. On this 'the Working Party noted that there was nothing in the Agreement, or in the rules for tariff negotiations which has been used in the past, to prevent governments from adopting any formula they might choose, and therefore considered that there was no need for the CONTRACTING PARTIES to make any recommendation in this matter' [GATT, BISD, Third Supplement, p. 22]. No further attempt has been made to give greater definition to the manner in which reciprocity is to be measured and it has been left to each county to develop its own yardsticks (Hoda, 2001, p. 53).

country is an exporter of the final good, then a reduction in τ_1^H raises p_1^H and therefore moves this price in the direction of the efficient level; but if the home country is an importer of the final good, then a reduction in τ_1^H reduces p_1^H and moves this price *further away* from the efficient level. Whether the resulting first-order loss in joint surplus can be made up with additions to joint surplus associated with the other elements of the reciprocal liberalization depends on circumstances, but with respect to the impact on p_1^H it is clear that reducing τ_1^H (reciprocally or otherwise) reduces joint surplus when the home country is an importer of the final good.

Hence, in the presence of offshoring our earlier answer to the first question above must at a minimum be qualified. And our earlier answer to the second question is overturned in this environment, for the same reason that it is overturned in the missing-instruments environment: as with missing instruments, in the presence of offshoring it is no longer generally the case that politically optimal policies are internationally efficient, because as Antràs and Staiger (2012a) show and as we have described earlier, there is a local price externality that may now persist at the political optimum. In fact, as we have discussed, Antràs and Staiger argue that in the presence of offshoring the political optimum will be inefficient unless government objectives correspond to national income maximization. And with the inefficiency of the political optimum in the presence of offshoring comes as well the conclusion that no point on the efficiency frontier is robust to renegotiation subject to reciprocity. We therefore conclude that the attractive features of reciprocity that hold in the context of the terms-of-trade theory and that we have emphasized above may be substantially diminished in environments where offshoring is present.

Finally, notice that at one level all the models we have reviewed thus far share a common perspective on the purpose of a trade agreement: the trade agreement exists to address an international externality that is associated with unilateral policy choices. As we have described, regarding their stance on the question of purpose, where these models sometimes differ is in the form that the externality takes. The "commitment theory" of trade agreements is in this respect quite different, in that the central role for an international externality is absent. According to the commitment theory governments value trade agreements as a way to tie their hands (make commitments) against their own lobbies and citizens.[cd] Does reciprocity look attractive when viewed from the perspective of the commitment theory?

To answer this question, we begin by observing that as an international externality plays no fundamental role in the commitment theory, it is natural that the theory would

[cd] Expressions of the commitment theory of trade agreements can be found in a variety of early papers (see, for example, Carmichael, 1987; Staiger and Tabellini, 1987; Lapan, 1988; Matsuyama, 1990; Tornell, 1991; and Brainard, 1994), but Maggi and Rodriguez-Clare's (1998) extension of the lobbying model of Grossman and Helpman (1994) provides what has become the workhorse model of the commitment theory in the trade agreements literature.

predict no *international* inefficiencies associated with the unilateral policy choices of governments. Rather, in the commitment theory the inefficiencies associated with unilateral policy setting are *domestic* in nature: they reflect domestic distortions that are created when governments have incentives to surprise domestic private actors (producers and/or consumers) with unexpected policy intervention, and lack the ability to precommit on their own not to engage in such behavior. In the context of trade policy the incentive to surprise is especially likely to be present, owing to the second-best nature of trade-policy intervention and the incentive governments have to bring such intervention closer to the first best through the element of surprise (see, for example, Staiger and Tabellini, 1989). The point can be seen clearly in the case where a government uses a tariff to address a domestic consumption distortion, say, a negative externality (that does not cross borders) associated with domestic consumption. As a tariff is a combination production subsidy and consumption tax, it is a second-best instrument for addressing the consumption distortion; and it is for this reason that the government would have an incentive to surprise, announcing a policy of nonintervention until domestic production decisions had been made, and then following through with a tariff once production decisions are sunk so as to discourage consumption. In equilibrium domestic producers would not likely be fooled, and as a result the government might be better off if it could commit not to use the tariff at all. And if it cannot manage this commitment on its own, a trade agreement may serve as a useful external commitment device.

It should be fairly clear from this discussion that where governments seek to make commitments to their own private sectors through trade agreements, the attractive features of reciprocity that we have emphasized above in the context of the terms-of-trade theory are likely to lose their luster. This can be most directly appreciated by considering the case of a truly small country that seeks trade agreements as a way to make commitments to its private sectors. For such a country, reciprocity has no bite whatsoever, because by definition the trade policy choices of this country cannot alter the terms of trade; and so, if this country were allowed to renegotiate out of its commitments subject to reciprocity, its commitments would become meaningless. Put slightly differently, according to the logic of the commitment theory, governments care about their own tariff commitments but have no particular reason to care about the tariff commitments of their trading partners, whether those commitments are reciprocal or not. The one potential caveat to this statement arises in regard to how the commitments will be enforced. As trade agreements must generally be self-enforcing, and as commitments have no value if they are not enforced, it is possible that some form of reciprocity would be desired even by governments who saw the trade agreement as valuable to themselves only in so far as the agreement helped them to make commitments to their own private sectors. The logic is that only with reciprocity in some form would these governments (and their private sectors) expect that the commitments they undertook in a trade agreement might actually be enforced. We see this logic as a possible route along which

reciprocity might be shown to play an important role in the commitment theory of trade agreements, and in that light as an interesting avenue for further research.

4. NONDISCRIMINATION

We have focused thus far on two-country modeling environments, but real-world trade agreements operate in a multicountry setting. With multiple countries, an important design feature of a trade agreement is whether it allows its member countries to adopt discriminatory tariffs against each other. In the GATT/WTO, such discriminatory tariff behavior is discouraged by the most-favored-nation (MFN) principle contained in GATT Article I, which prohibits any WTO member from applying its tariff on imports of a given good (a "like product") in a way that discriminates across exporters from different WTO member countries (or at a level that is higher than it applies to imports of the good from nonmember countries).[ce] Exceptions to the MFN principle are provided in GATT/WTO rules, however, and can be invoked by GATT/WTO member governments in certain circumstances. The most important exception to MFN is contained in GATT Article XXIV, which allows GATT/WTO members to form preferential trade agreements provided that the members of such agreements eliminate tariffs on "substantially all trade" between them. The coexistence of the GATT/WTO on the one hand, which is built on the foundation of the MFN principle, and on the other hand preferential trade agreements, of which there are now nearly 400 in force, raises the question: What are the advantages and disadvantages of the MFN principle as a design feature of a trade agreement? This is the question we address in this section.

4.1 A Three-Country Model, Discriminatory Tariffs, and MFN

To proceed, we begin by describing a three-country extension of the two-country two-good general equilibrium trade model developed in Section 2.[cf] In the three-country extension, the home country now imports good x from two foreign countries who for simplicity do not trade with each other, and exports good y to each of them. We denote foreign-country 1 and foreign-country 2 variables with the superscripts "$*1$" and "$*2$," respectively. Each foreign country can impose a tariff on its imports of good y from the home country, and we represent the tariff of foreign country i by τ^{*i}. The home country can set tariffs on its imports of good x from the two foreign countries; we represent the home-country tariff on imports from foreign country i by τ^{i}. An

[ce] In addition to the MFN principle, the National Treatment principle, contained in GATT Article III and prohibiting discriminatory treatment against foreign produced goods once they have cleared customs, is the other major leg of the nondiscrimination principle in the GATT/WTO.

[cf] Bagwell and Staiger (1999, 2005a, 2010b) provide details of this model and develop most of the themes we describe here. See also Bagwell et al. (2015).

important observation is that there can only be one local price in the home economy, and the pricing relationships $p = \tau^1 p^{w1}$ and $p = \tau^2 p^{w2}$ therefore imply $p^{w1} \neq p^{w2}$ whenever $\tau^1 \neq \tau^2$. That is, if the home country applies *discriminatory* tariffs $\tau^1 \neq \tau^2$ against the imports from foreign countries 1 and 2, then separate equilibrium world prices $\widetilde{p}^{w1}(\tau^1, \tau^2, \tau^{*1}, \tau^{*2})$ and $\widetilde{p}^{w2}(\tau^1, \tau^2, \tau^{*1}, \tau^{*2})$ apply to its trade with foreign countries 1 and 2, respectively. Alternatively, if the home country applies a *nondiscriminatory* (MFN) tariff $\tau^1 = \tau^2 \equiv \tau$ against the imports from foreign countries 1 and 2, then a single equilibrium world price $\widetilde{p}^{w}(\tau, \tau^{*1}, \tau^{*2})$ applies to its trade with both foreign countries.

4.1.1 Discriminatory Tariffs

Let us start with discriminatory tariffs. Owing to our assumption that each foreign country trades only with the home country, it is clear that the objectives of the foreign governments can be represented in this three-country model in a completely analogous fashion to that of the two-country model: with discriminatory home-country tariffs, foreign government welfares can be written as $W^{*1}(p^{*1}, \widetilde{p}^{w1})$ and $W^{*2}(p^{*2}, \widetilde{p}^{w2})$ with $W^{*1}_{\widetilde{p}^{w1}} > 0$ and $W^{*2}_{\widetilde{p}^{w2}} > 0$. But discriminatory tariffs complicate the expression for the welfare of the home government relative to that in the two-country model. In particular, we cannot simply express home government welfare as a function of the home local price and the two world prices, because the home government is not indifferent to the foreign source of its imports of x when p^{w1} and p^{w2} are not equal: the home government would prefer to have more of its imports of x coming from the foreign country with which its tariff is higher and its terms of trade more favorable (p^{wi} lower), and the share of its imports coming from each foreign source depends on the local prices in each foreign country. For this reason, as Bagwell and Staiger (1999) show, when the home government adopts discriminatory tariffs its welfare can now be represented by $W(p, T)$ with $W_T < 0$, where T is the home country's multilateral terms of trade defined as a trade-weighted average of the bilateral world prices with each of its trading partners and is therefore a function of both world prices *and* foreign local prices.

Hence, in the presence of discriminatory tariffs the channels through which externalities travel extend beyond world prices to include local prices. A number of conclusions relevant to trade agreements that permit tariff discrimination then follow. First, as Bagwell and Staiger (1999) show, these local price externalities are not neutralized at the political optimum, and so politically optimal tariffs are not efficient in the presence of discriminatory tariffs. As can be expected from our discussion in the previous section, this implies in turn that the attractive features of reciprocity that we emphasized above fail to survive in a multicountry world with discriminatory tariffs. Second, as Bagwell and Staiger (2005b) show, trade agreements that permit discriminatory tariffs are susceptible to an extreme form of bilateral opportunism through "concession erosion": beginning from any point on the efficiency frontier, the home government and the government of either of its trading partners can gain from a bilateral agreement in which each agrees

to lower its tariff against the other and worsen the terms of trade of the third country, converting the third-country loss into their own gain.[cg] This suggests that trade agreements that allow for discriminatory tariffs may have particular difficulty reaching and maintaining *any* position on the efficiency frontier.

4.1.2 MFN

Next consider the impact of the MFN principle, which in this setting requires that the home government apply a common tariff level $\tau^1 = \tau^2 \equiv \tau$ to the imports of x, regardless of whether these imports originate from foreign country 1 or 2. As we have observed, this implies that a single equilibrium world price $\widetilde{p}^w(\tau, \tau^{*1}, \tau^{*2})$ must then prevail. But then, in complete analogy to the two-country model, the objectives of the three governments may be written as $W(p, \widetilde{p}^w)$, $W^{*1}(p^{*1}, \widetilde{p}^w)$ and $W^{*2}(p^{*2}, \widetilde{p}^w)$, with $W_{\widetilde{p}^w} < 0$, $W^{*1}_{\widetilde{p}^{w1}} > 0$ and $W^{*2}_{\widetilde{p}^{w2}} > 0$.

Evidently, in a multicountry environment, the MFN principle ensures that the international externality continues to exhibit the same structure as in the simpler two-country setting. And as a result, in the company of MFN the attractive properties of reciprocity described in the previous section extend to the multicountry setting as well. In particular, under the MFN principle and beginning from noncooperative tariffs, each country can gain from reciprocal liberalization that does not go too far; and MFN politically optimal tariffs (which are efficient) are the only efficient tariffs that are robust to renegotiation under the reciprocity rule, suggesting that, once achieved, the political optimum can be viewed as a particularly robust and stable bargaining outcome of GATT tariff negotiations.[ch]

4.2 MFN Plus Reciprocity

The MFN principle can be understood to offer a further advantage when it is joined with reciprocity in a multicountry world. As we next describe, together reciprocity and MFN can neutralize third-party externalities of bilateral tariff bargaining.

[cg] With previous concessions exchanged to position the three countries initially on the efficiency frontier, and with the deterioration in foreign country 2's terms of trade synonymous with a reduction in its access to the home-country market, it is natural to use the term "concession erosion" to describe the impact on foreign country 2 of the bilateral bargain between the home country and foreign county 1.

[ch] It is also straightforward to extend to the multicountry MFN setting our earlier finding that, if governments wish to adopt a common terms of exchange of market access for all countries, the adding-up constraint imposed by market clearing requires that they adopt the one-for-one terms of exchange embodied in GATT's reciprocity principle. A different view of the benefits of the MFN principle than we have emphasized in the text is offered by McCalman (2002). In McCalman's model private information plays a central role, and if the number of small countries is sufficiently great, MFN helps to diminish the ability of a large country to hold an agreement hostage and extract rents from its small trading partners, thereby enhancing both global efficiency and the payoffs of small countries. Other implications of the MFN principle are explored in Ludema (1991), Choi (1995), and Saggi (2004).

4.2.1 The Bilateral Opportunism Problem and the Free-Rider Problem

To appreciate the additional advantage of MFN when it is combined with reciprocity, we first return to the bilateral opportunism problem raised above in the presence of discriminatory tariffs, and note that the MFN restriction by itself mitigates against this problem (as Schwartz and Sykes, 1997, point out) but does not completely prevent it (as emphasized by Bagwell and Staiger, 2005b). Suppose for example that the home country and foreign country 1 engage in a bilateral bargain to cut their tariffs, with the home-country tariff constrained to abide by MFN. If foreign country 1 agrees to cut its tariff unilaterally, this will worsen its terms of trade, which under MFN is to say it will also worsen the terms of trade of foreign country 2 (who experiences concession erosion) while improving the terms of trade of the home country. Under this agreement, the bilateral bargain between the home country and foreign country 1 would impose a negative externality on Foreign country 2. Of course, unless foreign country 1 begins the negotiations with its tariff above its own reaction curve, it would lose from such a unilateral tariff cut, but it *could* gain from the bargain if the home country agrees to cut its tariff as well, and it *might* gain even if the agreed changes to the two tariffs together still result in a deterioration of the foreign (country 1 and country 2) terms of trade. Bagwell and Staiger show that the MFN restriction would prevent the home country and foreign country 1 from jointly gaining in such a bilateral bargain beginning from some points on the efficiency frontier (the MFN political optimum included among them), thereby confirming the position of Schwartz and Sykes that MFN can help protect against concession erosion; but Bagwell and Staiger show that there are also points on the efficiency frontier from which, even under the MFN restriction, the home country and foreign country 1 *can* jointly gain from engineering concession erosion and the associated terms-of-trade deterioration suffered by foreign country 2.

Hence, the negative third-party externality problem associated with concession erosion is mitigated under the MFN restriction, but only partly so.[ci] At the same time, it is also true that MFN introduces a new concern associated with a *positive* externality to third parties: the free rider problem. This is the familiar weakness of the MFN principle, which in our three-country model requires that any tariff cut given by the home country to one foreign country must be automatically (and unconditionally) extended as well to the other foreign country. In the context of the bilateral bargain between the home country and foreign country 1 that we considered just above, the free-rider problem associated with MFN arises when the home country agrees to cut its tariff, because in so doing the home country is then worsening its own terms of trade and conferring a terms-of-trade

[ci] And as Bagwell and Staiger (2010a, note 10) point out, this partial mitigation is further weakened if international transfers are possible, providing a potential reason to discourage the use of transfers in the GATT/WTO system.

benefit to foreign country 1, which under MFN is to say it is also conferring a terms-of-trade benefit to foreign country 2, who "free rides" on the additional market access that the home country's tariff cut implies. The free-rider problem is typically viewed as the Achilles heel of the MFN principle.

4.2.2 MFN Plus Reciprocity

Notice, however, that what we describe above is in the first instance the possibility that a bilateral MFN tariff bargain might impose a *negative* externality on third parties; there the negative externality arises when foreign country 1 cuts its tariff on imports of good y in a bilateral bargain with the home country, and the externality is transmitted to competing importers of good y (foreign country 2). And what we describe in the second instance is a *positive* third-party externality that arises when the home country cuts its tariff on imports of good x in a bilateral bargain with foreign country 1 and is transmitted to competing exporters of good x (foreign country 2). Suppose, then, that the home country and foreign country 1 were to engage in a bilateral MFN tariff bargain that cut the tariff of foreign country 1 and the tariff of the home country in a way that just *balanced* these two opposing third-party externalities; in principle, the home country and foreign country 1 could then neutralize the third-party externality of their bilateral tariff bargain. As Bagwell and Staiger (2005b) show, this balance is precisely what GATT's principle of reciprocity achieves in a multicountry MFN world, because the balance requires nothing more than ensuring that the terms of trade does not move in either direction as a result of the bilateral bargain, and that is what MFN tariff changes that abide by reciprocity deliver.

It bears emphasis that, following a bilateral reciprocal MFN tariff bargain between the home country and foreign country 1, exporters from foreign country 2 experience a reduced MFN tariff from the home country and yet do not enjoy any increase in their export volume. How can this be? The reason is that the exports from foreign country 2 must compete for sales in the domestic market with the exports from foreign country 1, and exports from foreign country 1 are stimulated by the negotiated reduction in foreign country 1's import tariff (Lerner symmetry).[cj] Put another way, what we are describing here is the possibility that tariff adjustments in one country (the home country) that, in isolation, would have altered the local prices in a second country (foreign country 2) and as a consequence triggered changes in that country's export volumes, might be matched by the tariff adjustments of a third country (foreign country 1) in a way that ultimately left the local prices of the second country, and hence its export volumes, unchanged. This reasoning is admittedly subtle, and there is as yet scant evidence available on its empirical

[cj] In this light, it is also easy to see that in a many-good setting where reciprocity could lead to changes in individual world prices but not the overall terms of trade, the third country could experience changes in the export volumes of individual goods but not in its overall export volume.

relevance.[ck] There is, however, ample evidence on the empirical relevance of Lerner symmetry, and the point we describe here is simply an application of Lerner symmetry to the context of bilateral MFN tariff bargaining.[cl]

Finally, we note that Ossa (2014) provides an important qualification to the claim that reciprocity and MFN can neutralize the third-party externalities of bilateral tariff bargaining, arguing that the existence of differentiated products can interfere with this claim. Intuitively, in a setting with differentiated products, even if it imposes an MFN tariff a country can trade at distinct world prices for (differentiated) products exported from distinct foreign sources, contrary to the homogeneous-good case; and in this setting, negotiating reciprocal tariff changes with one foreign partner that leave unchanged the terms of trade with that partner will generally alter (and as Ossa demonstrates, in fact improve) the terms of trade with the other foreign partner.[cm] Of course, with product differentiation the definition of "like products" takes center stage in the application of the MFN principle: At what point do differentiated products cease to become like products and hence not demand a common tariff treatment under MFN? The answer to this question is especially important here, because as Ossa demonstrates and as is intuitively clear, the magnitude of the third-party externality imposed by bilateral bargains that respect reciprocity and MFN is proportional to the degree of product differentiation within the scope of the like-product determination over which the MFN restriction applies; in the GATT/WTO this scope is typically constrained to fall within narrow product classifications. And as Ossa observes, for his quantification exercises he must impose MFN at the level of broad industry categories. Nevertheless, with this caveat noted Ossa reports

[ck] But we note here that, though with the purpose of making a different point, the Haberler Report (GATT, 1958, pp. 95–96) includes discussions that mirror a key part of the reasoning we sketch above, namely, that multiple tariff adjustments can have offsetting impacts on local prices with no consequence for trade volumes.

[cl] See Bagwell and Staiger (2014) for further discussion of empirical studies reporting evidence that a country's own tariff cuts stimulate its exports. Two recent papers provide evidence on the extent of the free-rider problem in GATT/WTO tariff negotiations more generally, though neither provides specific evidence on the contribution of reciprocity in mitigating the problem. Ludema and Mayda (2013) report evidence of significant free-riding effects in the pattern of tariff commitments implemented as a result of the 1986–94 Uruguay Round of GATT negotiations, while Bagwell et al. (2015) fail to find evidence of a free-rider problem in their analysis of the detailed bargaining records of the 1950–51 Torquay Round of GATT negotiations.

[cm] In this sense, Ossa's (2014) finding shares similarities with the finding of Bagwell and Staiger (2005b), that with discriminatory tariffs across trading partners—and hence multiple world prices for the same good—reciprocity does not eliminate third-party externalities. A difference, though, is that with discriminatory tariffs (and homogeneous goods) Bagwell and Staiger show that reciprocal tariff liberalization between two parties will hurt the third party via a terms-of-trade deterioration, while Ossa shows with MFN tariffs and differentiated products that the third party benefits via a terms-of-trade improvement. See also Suwanprasert (2014), who reports an analogous result to that of Ossa (2014) in the firm-delocation model of Ossa (2011).

that in the presence of differentiated products the third-party externality associated with MFN can be substantial, even when negotiations conform closely to reciprocity.[cn]

4.3 Multilateral Reciprocity

Overall, the preceding discussion suggests an important insight: broadly speaking, the MFN principle permits the liberalizing force of reciprocity to be harnessed in an essentially bilateral manner even in a multicountry world. Indeed, in the context of reciprocity's implications for the GATT/WTO bargaining process discussed in the earlier two-country setting, it can be shown that when that setting is extended to a multicountry world, reciprocity and MFN together work to eliminate bargaining externalities across bargaining pairs, and at the same time induce truth-telling on the part of governments, and in this way convert a potentially complex multilateral bargaining problem with private information into a comparatively simple set of full-information bilateral bargains (albeit with difficulties encountered in the face of asymmetries as we described earlier).[co]

But this observation raises a further question: If MFN and reciprocity can together essentially "decentralize" the tariff problem in a multicountry world into a collection of bilateral bargains, why, then, is there a need for GATT/WTO multilateral rounds? Why not simply let countries negotiate a web of bilateral reciprocal MFN tariff agreements on their own, much as was the practice of the United States during the 1930s under the US Reciprocal Trade Agreements Act (RTAA)?

One answer to this question is that, at least in the later GATT/WTO rounds, writing and elaborating on agreed common codes of behavior has been an important function, and for this the multilateral features of the negotiations are no doubt crucial. A second answer to this question is provided by Maggi (1999). Maggi demonstrates that, even in a world without externalities across bilateral relationships, multilateral tariff bargaining yields a Pareto-superior outcome to bilateral tariff bargaining whenever there are bilateral imbalances of power across bargaining pairs. As Maggi explains, bilateral bargains generate outcomes that are biased in favor of the "strong" country in each bilateral relationship, and unless countries can make international lump sum transfers this results in an inefficiency at the global level that multilateral tariff bargaining can avoid.

Here we suggest a third answer to the question why multilateral rounds of negotiation might be desirable: the GATT/WTO multilateral rounds relax the requirement of

[cn] In their analysis of the Torquay bargaining records, Bagwell et al. (2015) report evidence which they interpret as consistent with Ossa's (2014) result that third-party externalities should be larger where products are more differentiated. In addition, Ossa (2014) reports results from counterfactuals that are designed to illuminate how the MFN principle by itself impacts GATT/WTO bargaining outcomes. Again as Ossa notes, imposing MFN at the broad industry categories used in his quantification exercise is problematic, but with this caveat he reports that the MFN restriction appears to have little quantitative impact on bargaining outcomes.

[co] See Bagwell and Staiger (1999, 2010b) and Bagwell et al. (2015).

bilateral reciprocity and allow countries to seek instead only *multilateral reciprocity* in their bargaining outcomes. And as we explain, in a multicountry MFN world, multilateral reciprocity is sufficient to deliver all of the properties that we have above attributed to the pairing of reciprocity with MFN.[cp]

To see the point, it is helpful to return to the three-country two-good general equilibrium model described earlier, with the home government restricted to an MFN tariff. As we observed, this implies that there is a single equilibrium world price $\widetilde{p}^w(\tau, \tau^{*1}, \tau^{*2})$, and under standard conditions $\widetilde{p}^w$ is decreasing in the home tariff τ and increasing in each of the foreign tariffs τ^{*1} and τ^{*2}. Let us suppose, then, that the tariff reductions $\Delta\tau$, $\Delta\tau^{*1}$ and $\Delta\tau^{*2}$ together satisfy reciprocity, in the sense that together they leave $\widetilde{p}^w$ unchanged.

One way to structure these tariff changes is to proceed sequentially through the bilaterals and demand bilateral reciprocity at every stage. Under this approach, the home country could split its tariff reduction $\Delta\tau$ into two parts: a $\Delta\tau_1$ that was matched with $\Delta\tau^{*1}$ so that $\widetilde{p}^w$ remained unchanged under the tariff reductions $\Delta\tau_1$ and $\Delta\tau^{*1}$; and a further $\Delta\tau_2$ that was matched with $\Delta\tau^{*2}$ so that $\widetilde{p}^w$ remained unchanged under the tariff reductions $\Delta\tau_2$ and $\Delta\tau^{*2}$ (and by construction, $\Delta\tau_1 + \Delta\tau_2 = \Delta\tau$). "Split concessions" of this kind were a common tactic used by the United States to maintain reciprocity in each of its bilaterals under the RTAA (see Beckett, 1941).

An alternative way to structure these tariff changes, however, is to not worry about achieving bilateral reciprocity in each bilateral bargain, as long as overall *multilateral* reciprocity is achieved once the direct-plus-indirect impacts of all the bilateral bargains on each country's trade are taken into account. Under this approach, in its bilateral with foreign country 1 the home country could offer its entire tariff cut $\Delta\tau$ in exchange for the tariff cut $\Delta\tau^{*1}$ from foreign country 1, creating a positive spillover for foreign country 2 from this bilateral; and in its bilateral with foreign country 2, the home country could then ask for the tariff cut $\Delta\tau^{*2}$ and offer nothing directly in return, with the tariff cut $\Delta\tau^{*2}$ serving as compensation for the indirect benefit that foreign country 2 would be receiving from the bilateral between home and foreign country 1. Neither of these bilaterals satisfies bilateral reciprocity, but each country nevertheless achieves reciprocity—and $\widetilde{p}^w$ is left unchanged—once the direct and indirect effects of all the tariff cutting agreed in the round are added up.

It is not hard to see that the second approach would be facilitated by a multilateral round of bilateral negotiations in which all participants in the bilaterals could assess both what they were obtaining directly in their own bilaterals and also what they would be obtaining indirectly from the bilaterals of other pairs of countries. In the simple model here, there is nothing to gain from relaxing the restriction of bilateral reciprocity and replacing it with the weaker requirement of multilateral reciprocity. But it is easy to see with more complicated trade patterns that the relaxation of the constraint from

[cp] This is a point emphasized by Bagwell et al. (2015).

one of bilateral reciprocity to one of multilateral reciprocity could make possible some deals that would otherwise not be possible.[cq]

In a four-country version of this model with one home and three foreign countries, Bagwell et al. (2015) show that if the home country and foreign countries 1 and 2 engage in negotiations that conform to MFN and multilateral reciprocity, then foreign country 3 is unaffected by their bargain. And with this result they also establish that multilateral reciprocity is sufficient to deliver all of the properties attributed above to the pairing of reciprocity with MFN. Bagwell et al. then use these properties to guide their empirical analysis of GATT bargaining records from the Torquay Round, and find evidence consistent with a number of themes developed above. But for our purposes here their findings concerning multilateral vs bilateral reciprocity are the most relevant. Exploiting the "quasi experiment" created by the breakdown of the United States–United Kingdom bilateral in the middle of the Torquay Round, Bagwell et al. report evidence of rebalancing in the offers and counteroffers made by the United States and its other bilateral bargaining partners subsequent to the news of this breakdown that would not have been required to maintain reciprocity if countries had been achieving bilateral reciprocity in each of their bargains all along. These findings support the view that the bilateral reciprocity restriction that might otherwise have constrained negotiations was indeed relaxed in the GATT multilateral bargaining forum, where the weaker restriction of multilateral reciprocity could instead be achieved.[cr]

5. BINDINGS AND OVERHANG

Governments in the GATT/WTO negotiate bound tariff rates, or tariff caps, rather than exact tariff levels. When a government agrees to a tariff cap for some product, the government is permitted under GATT/WTO rules to set any (nondiscriminatory) tariff below that cap but is not allowed to apply a tariff above the cap unless certain contingencies arise.[cs] Binding overhang occurs when the applied tariff is below the bound level.

[cq] Indeed, the relaxation of this constraint was seen by early GATT practitioners as a key innovation of GATT relative to the RTAA (Interim Commission for the ITO, 1949). Notice, too, that our point here is distinct from that made by Maggi (1999), as our point relates to the ability to balance spillovers across bilaterals in the presence of MFN whereas Maggi's point addresses power asymmetries in bilaterals with no spillovers and no essential role for MFN.

[cr] A remaining question is why, given the multilateral bargaining forum, the organizing principle of GATT/WTO tariff negotiations remained focused on bilaterals. In fact, in several of the GATT Rounds (the Kennedy Round and the Tokyo Round) countries first negotiated on a multilateral basis over a tariff cutting rule (or rules) to be adopted by all, and the bilaterals then reemerged in these rounds as a way of introducing exceptions to the agreed rules and to establish and maintain reciprocity. In the now-stalled WTO Doha Round, a mixture of these approaches was being utilized. Especially with the recent declassification of the GATT bargaining records for many of the GATT rounds, we see the choice of bargaining protocols to be an important topic for future theoretical and empirical research.

[cs] Tariff caps play a central role in GATT/WTO design. This role is highlighted, for example, in the World Trade Report 2009 (World Trade Organization, 2009, p. 105), which argues that the "concept of a tariff binding—ie, committing not to increases a duty beyond an agreed level—is at the heart of the multilateral trading system."

In practice, the extent of binding overhang varies markedly across countries and products, but is a prominent feature of tariff policies for many WTO members.[ct] Thus, GATT/WTO rules allow for substantial "downward flexibility" but impose significant constraints on "upward flexibility."

Does the GATT/WTO tariff-cap approach make sense in light of the trade-agreement models described in Section 2? The terms-of-trade theory of trade agreements clearly suggests that an effective trade agreement must include some constraints against opportunistic tariff hikes. A tariff cap is one such constraint. From the perspective of this theory, we can also appreciate that a well-designed trade agreement might provide substantial downward flexibility; after all, if a government applies a (nondiscriminatory) tariff that is below the bound level for a given product, then the trading partners that export this product enjoy a terms-of-trade gain. A further consideration is that the nature of tariff flexibility may impact the negotiated baseline tariff; for example, a higher baseline tariff may be more attractive in the presence of downward flexibility. To address these and other considerations and to thereby assess the optimality of the GATT/WTO tariff-cap approach, we require a model of tariff negotiations in which shocks occur so that flexibility would sometimes be exercised if permitted.

Bagwell and Staiger (2005a) analyze tariff caps and binding overhang in the context of a model in which governments are exposed to preference shocks in the form of political pressure. They posit that at the time of negotiation governments face at least some uncertainty about the political pressures that they might later face. After the trade agreement is negotiated, governments privately observe their respective political pressure realizations and select their preferred import tariffs from those that are permitted by the trade agreement. Bagwell and Staiger develop their analysis in the context of the linear-quadratic partial-equilibrium model presented in Section 2.2.1, under the assumptions that only import tariffs are selected (ie, export policies are exogenously set at free trade), the political-economy weights attached to export interests are unitary, and the political-economy weights that governments attach to import-competing interests are independently drawn from the interval $[\underline{\gamma},\overline{\gamma}]$ according to the continuously differentiable distribution function $F(\gamma)$ and where $\underline{\gamma}=1<\overline{\gamma}<7/4$.[cu]

Given the separable structure of the linear-quadratic model and the absence of export policies, the tariff policy of one government has no direct interaction with the tariff selection of the other government. We therefore focus on the home-country import good,

[ct] See, for example, WTO (2009, p. xix) for a related description. Recent empirical work includes Bacchetta and Piermartini (2011) and Beshkar et al. (2015).

[cu] Thus, in terms of the model presented in Section 2.2.1, Bagwell and Staiger (2005a) assume that $\gamma_e=\gamma_e^*=1$ while γ_m and γ_m^* are independently determined according to the distribution function $F(\gamma)$. Recall now the assumption that $5/8+9\gamma_e/8>\gamma_m$ for the linear-quadratic model in Section 2.2.1, which ensures that trade volume is positive at Nash (and efficient) tariffs. In the present context, with $\gamma_e=1$, this assumption requires that $7/4>\gamma_m$, which explains the assumption that $\overline{\gamma}<7/4$.

which we denote as good x in Section 2.2.1.[cv] To simplify our presentation, we drop the good x subscript in our notation here.

Consider then the import tariff policy of the home country. For the linear-quadratic model, the optimal tariff for the government of the home country is the same as its Nash tariff.[cw] As Bagwell and Staiger (2005a) confirm, the *Nash tariff function* for the linear-quadratic model takes the form $\tau^N(\gamma) = (8\gamma - 5)/[4(17 - 2\gamma)]$, where γ is the political-economic welfare weight attached to profit in the import-competing industry. As illustrated in Fig. 3 (adapted from Bagwell and Staiger, 2005a), the Nash tariff function is strictly increasing. This simply reflects the fact that a home-country government with a greater welfare weight on import-competing interests derives greater benefit from the higher local price that a tariff hike implies. Similarly, we let $\tau^E(\gamma)$ denote the import tariff for the home-country government that maximizes joint government welfare, $W + W^*$. Bagwell and Staiger show that $\tau^E(\gamma) = 4(\gamma - 1)/(25 - 4\gamma)$. This function is strictly increasing as well, and it is also strictly lower than the Nash tariff: $\tau^E(\gamma) < \tau^N(\gamma)$. The difference between the two tariffs is attributable to the terms-of-trade externality. A final feature illustrated in Fig. 3 is that the difference between $\tau^N(\gamma)$ and $\tau^E(\gamma)$ shrinks as γ rises. The intuition is that the incentive to manipulate the terms of trade diminishes at higher values for γ, since $\tau^N(\gamma)$ and $\tau^E(\gamma)$ then take higher values and thus generate lower trade volumes. Of course, $\tau^N(\gamma)$ and $\tau^E(\gamma)$ also characterize the tariff policies for the foreign government, when γ is interpreted as the welfare weight that the foreign government attaches to its import-competing industry (for good y).

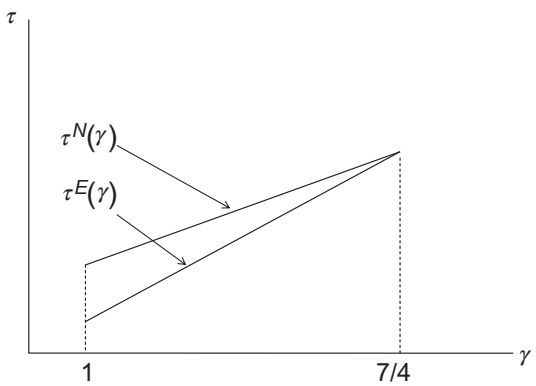

Fig. 3 Nash and efficient domestic tariff levels.

[cv] Even though the model has a separable structure, in the private-information setting under consideration here, gains may exist from "linking" tariff decisions over time or across products. We return to this issue below.

[cw] Amador and Bagwell (2013) refer to this tariff as the flexible tariff, since it is the tariff that a government would select if it were granted full flexibility under the trade agreement.

Since we have ruled out export policies, governments do not have sufficient trade-policy instruments with which to effect lump-sum transfers. The relationship between efficient tariffs and the function $\tau^E(\gamma)$ thus requires some additional discussion. We make two points. First, if governments have available an ex ante (noncontingent) lump-sum transfer instrument, then they achieve efficiency by maximizing $W + W^*$, and so $\tau^E(\gamma)$ then indeed defines the efficient tariff function. Second, if such an instrument is not available, then the Pareto frontier includes other tariff functions as well; however, in the symmetric environment under consideration here, a focal point on the efficiency frontier is arguably where governments use the same tariff function, $\tau^E(\gamma)$. We therefore refer below to the tariff function $\tau^E(\gamma)$ as the *efficient tariff function*, and we evaluate a trade agreement relative to the objective of maximizing the expected value of $W + W^*$. The efficient tariff function is the first-best function for this objective, since it maximizes the objective pointwise.

Unfortunately, however, the efficient tariff function is not feasible in this private-information setting, since it fails to be incentive compatible. To see the issue, we may refer again to Fig. 3 for a given value of γ, the home-country government prefers $\tau^N(\gamma)$ to $\tau^E(\gamma)$. If the governments were to attempt to implement the efficient tariff function, then the home-country government with true political pressure γ would have incentive to claim that its type is $\gamma' > \gamma$ where $\tau^E(\gamma') = \tau^N(\gamma)$. Thus, if governments are unable to observe and write verifiable contracts that condition on political pressure realizations, and if governments do not have available contingent transfer functions with which to make the efficient tariff function incentive compatible, then a trade agreement cannot deliver the efficient tariff function.[cx]

Bagwell and Staiger (2005a) compare a *strong tariff binding*, whereby a government must select the same tariff for all political-pressure realizations, with a *weak tariff binding*, whereby a government can apply any tariff so long as the applied tariff does not exceed the bound tariff level (ie, the tariff cap). Given that political pressure enters welfare in a linear way, they find that the optimal strong tariff binding is at the tariff $\tau^E(E(\gamma))$, where $E(\gamma)$ is the expected value of political pressure. Imposing the assumption that $E(\gamma) > 5/4$ so that $\tau^E(E(\gamma)) > \tau^N(\underline{\gamma})$, they further show that the optimal weak binding, τ^w, strictly exceeds $\tau^E(E(\gamma))$ and corresponds to the efficient tariff level for the average political pressure among those types that are constrained by the binding (ie, among those higher types for which $\tau^N(\gamma) > \tau^w$). Intuitively, when a weak binding is used, lower types automatically satisfy the binding while selecting their Nash tariffs; thus, the bound rate is actually selected only by higher types, and so the optimal weak binding is efficient on average conditional on the realization of higher types.

[cx] As Bagwell and Staiger (2005a) discuss, since the efficient tariff function is strictly increasing, it could be implemented were lump-sum and contingent transfers feasible. In practice, however, state-contingent and monetary transfers are not a prominent feature of GATT/WTO practice.

For both the strong and weak tariff bindings, the induced applied tariff schedules are incentive compatible and thus correspond to a feasible trade agreement.[cy] Furthermore, and as Bagwell and Staiger (2005a) show, the optimal weak binding generates strictly greater expected joint welfare than does the optimal strong binding.[cz] In this sense, their analysis provides a rationalization for the GATT/WTO tariff-cap approach, while also providing an interpretation of binding overhang (by lower types for which $\tau^N(\gamma) < \tau^w$). At the same time, the analysis is not fully complete in that it does not establish that the trade agreement defined by the optimal weak binding offers greater expected joint welfare than does any other incentive-compatible trade agreement.

Amador and Bagwell (2013) develop a theory of optimal delegation that completes and extends this analysis. In a delegation problem, a principal selects a set of permissible actions from which a privately informed agent selects.[da] The principal faces a trade off when selecting this set, since the agent has superior information (which argues in favor of giving the agent substantial discretion) but has biased preferences (which argues in favor of limiting the agent's discretion). Amador and Bagwell characterize optimal delegation for a general family of preferences for the principal and agent, and they then apply their characterizations to the trade-agreement problem. Specifically, they show that the trade-agreement problem can be understood as a delegation game in which the "principal's" objective corresponds to the ex ante expected joint welfare of the two governments, the privately informed "agent's" objective corresponds to the ex post welfare of the government of the importing country, the state of the world about which the agent obtains private information corresponds to the level of political pressure that this government ultimately faces, and the set of permissible actions corresponds to the set of tariffs that are allowed under the trade agreement.

Formally, Amador and Bagwell (2013) analyze the following *delegation problem*:

$$\begin{aligned} &\max_{\pi:\Gamma\to\Pi} \int_{\Gamma} w(\gamma,\pi(\gamma))dF(\gamma) \text{ subject to} \\ &\gamma \in \arg\max_{\tilde{\gamma}\in\Gamma} [\gamma\pi(\tilde{\gamma}) + b(\pi(\tilde{\gamma}))] \text{ for all } \gamma \in \Gamma, \end{aligned} \tag{75}$$

where $\Gamma \equiv [\underline{\gamma}, \overline{\gamma}] \subset \Re$, F is continuous with an associated continuous and strictly positive density f, Π is an interval of the real line with nonempty interior where, without loss of generality, $\inf\Pi = 0$ and $\overline{\pi}$ is in the extended reals such that $\overline{\pi} = \sup\Pi$. Their maintained

[cy] The applied tariff schedule induced by the strong tariff binding is trivially incentive compatible, since all types apply the same tariff. The applied tariff schedule induced by the weak tariff binding entails lower types selecting their Nash tariffs and higher types pooling at the weak binding, τ^w. This applied tariff schedule is also incentive compatible, since each type applies the permitted tariff that is as close as possible to its Nash tariff.

[cz] To establish this point, suppose that governments were to use a weak binding where the cap is set at the optimal strong tariff binding, $\tau^E(E(\gamma))$. Given $\tau^N(\underline{\gamma}) < \tau^E(E(\gamma))$, it follows that the weak binding leads to strictly lower tariffs for lower types, with both governments strictly gaining as a result. The optimal weak binding is then an even better alternative for governments.

[da] See Holmstrom (1977) for an original statement and analysis of the delegation problem.

assumptions are that: (i) $w:\Gamma\times\Pi\to\Re$ is continuous on $\Gamma\times\Pi$; (ii) for all $\gamma_0\in\Gamma$, $w(\gamma_0,\cdot)$ is concave on Π and twice differentiable on $(0,\overline{\pi})$; (iii) $b:\Pi\to\Re$ is strictly concave on Π and twice differentiable on $(0,\overline{\pi})$; (iv) there exists a twice-differentiable function $\pi_f:\Gamma\to(0,\overline{\pi})$ such that, for all $\gamma_0\in\Gamma$, $\pi_f'(\gamma_0)>0$ and $\pi_f(\gamma_0)\in\arg\max_{\pi\in\Pi}[\gamma_0\pi(\gamma)+b(\pi(\gamma))]$; and (v) $w_\pi:\Gamma\times(0,\overline{\pi})\to\Re$ is continuous on $\Gamma\times(0,\overline{\pi})$.[db]

The objective of the principal is thus to maximize the expected value of the function w, where w depends on the state variable γ and the associated action, $\pi(\gamma)$. The agent privately observes the state γ and chooses over actions π to maximize $\gamma\pi + b(\pi)$. The principal, however, selects the set of permissible actions, and is thereby able to match states to actions provided that the incentive compatibility constraint is met (so that an agent that observes state γ does not prefer to choose a permissible action $\pi(\widetilde{\gamma})$ intended for a different state). Amador and Bagwell (2013) provide sufficient and necessary conditions under which an *interval allocation* solves the delegation problem defined in (75), where an interval allocation π is defined by bounds γ_L and γ_H with $\gamma_L,\gamma_H\in\Gamma$ and $\gamma_L<\gamma_H$ and satisfies $\pi(\gamma)=\pi_f(\gamma_L)$ for $\gamma\in[\underline{\gamma},\gamma_L]$, $\pi(\gamma)=\pi_f(\gamma)$ for $\gamma\in(\gamma_L,\gamma_H)$, and $\pi(\gamma)=\pi_f(\gamma_H)$ for $\gamma\in[\gamma_H,\overline{\gamma}]$. Thus, when the principal selects an interval allocation, an agent that observes an intermediate value of γ selects the optimal (ie, flexible) action $\pi_f(\gamma)$ while an agent that observes a higher (lower) value of γ selects the highest (lowest) permissible action. A *cap allocation* is an interval allocation for which $\gamma_L=\underline{\gamma}$.

A case of particular interest for trade-agreement applications occurs when $w(\gamma,\pi(\gamma))=v(\pi)+b(\pi)+\gamma\pi$. For example, consider again a partial equilibrium model of trade in which only import tariffs are available. For a given traded good, the interpretation here is that π is the profit that is delivered to the import-competing industry, with γ then indicating the welfare weight that the government of the importing country ultimately places on π. Since a higher import tariff results in a higher profit level in the import-competing industry, there is a one-to-one relationship between import tariffs and import-competing industry profit. The optimal import tariff, for example, is the unique tariff that delivers the corresponding optimal profit level, $\pi_f(\gamma)$. More generally, a trade agreement that specifies permissible import tariffs can be understood equivalently as specifying permissible levels of profit for the import-competing industry. We then may use $b(\pi)$ to capture the consumer surplus and tariff revenue in the importing country as a function of the delivered profit, so that the welfare function for the government of the importing country can be represented as $b(\pi)+\gamma\pi$. Finally, when the two governments design the trade agreement, they seek to maximize their ex ante expected welfare, $w(\gamma,\pi(\gamma))=v(\pi)+b(\pi)+\gamma\pi$, where the function $v(\pi)$ captures the impact of π on welfare in the exporting country. In other words, $v(\pi)$ represents export-industry profit and consumer surplus for the traded good in the exporting country, when these values are expressed as functions of the profit that is enjoyed on this good in the importing country.

[db] Amador and Bagwell (2013) also analyze and apply a version of the delegation problem that allows for a nonnegative money-burning variable.

For trade-agreement applications, cap allocations correspond to tariff caps and are thus of particular interest. Under the assumption that $v'(\pi_f(\gamma)) < 0$ for all $\gamma \in \Gamma$, Amador and Bagwell (2013) show that, within the family of cap allocations, the optimal value for γ_H is *interior* (ie, satisfies $\gamma_H \in (\underline{\gamma}, \overline{\gamma})$) provided that $v'(\pi_f(\underline{\gamma})) + E(\gamma) - \underline{\gamma} > 0$. The assumption that $v'(\pi_f(\gamma)) < 0$ is quite natural in trade-agreement applications, where a higher import tariff imposes a negative international externality on a country's trading partner. For example, in the linear-quadratic partial equilibrium model that Bagwell and Staiger (2005a) use, the assumption that $v'(\pi_f(\gamma)) < 0$ for all $\gamma \in \Gamma$ is satisfied, and the requirement that $v'(\pi_f(\underline{\gamma})) + E(\gamma) - \underline{\gamma} > 0$ holds given $\underline{\gamma} = 1$ if and only if $E(\gamma) \geq 5/4$.[dc] The tariff cap implied by the optimal cap allocation in this example corresponds to the optimal weak binding, τ^w, that Bagwell and Staiger (2005a) characterize.

For the case where $w(\gamma, \pi(\gamma)) = v(\pi) + b(\pi) + \gamma\pi$, Amador and Bagwell (2013) further provide a simple set of sufficient conditions under which the optimal cap allocation solves the delegation problem defined in (75) and is thus the optimal allocation among *all* incentive-compatible allocations. The sufficient conditions are those just mentioned—that $v'(\pi_f(\gamma)) < 0$ for all $\gamma \in \Gamma$ and $v'(\pi_f(\underline{\gamma})) + E(\gamma) - \underline{\gamma} > 0$—as well as two additional conditions. The additional conditions are that the density f is nondecreasing on Γ and that

$$\kappa \equiv \inf_{(\gamma,\pi)\in\Gamma\times\Pi} \left\{ \frac{w_{\pi\pi}(\gamma,\pi)}{b''(\pi)} \right\} \geq \frac{1}{2}. \tag{76}$$

The latter condition allows that v may be convex but ensures that the magnitude of any convexity in v is not too large relative to the magnitude of the concavity in b. The stated conditions are sufficient to ensure that the optimal cap allocation generates higher expected welfare for the principal than does any other incentive-compatible allocation.[dd]

As Amador and Bagwell (2013) show, these conditions can be easily applied to trade models. As a first application, they apply their analysis to the linear-quadratic partial equilibrium model studied by Bagwell and Staiger (2005a) and show that the optimal tariff-cap allocation is optimal in the full set of incentive-compatible allocations given $\underline{\gamma} = 1$ if $E(\gamma) \geq 5/4$ and f is nondecreasing. For this model, $\kappa = 2/3$ and so the condition that $\kappa \geq 1/2$ is automatically satisfied. Amador and Bagwell also show for this model that the monotonicity restriction on the density can be relaxed to allow for differentiable densities that are decreasing over ranges or even over the whole support, provided that the rate of decrease is not too great.[de] Thus, under the described conditions, the optimal weak

[dc] As Amador and Bagwell (2013) show, for the linear-quadratic model, $b(\pi) = (1/2)(-1 + 9\sqrt{\pi} - 17\pi)$ and $v(\pi) = (1/4)(2 - 6\sqrt{\pi} + 9\pi)$, where π is the profit in the import-competing industry and $\overline{\pi} = 1/9$ is the maximal profit obtained when the prohibitive tariff of 1/6 is applied.

[dd] Alternative incentive-compatible allocations must be nondecreasing and include a large family of allocations with points of discontinuity, where the allocation hurdles the flexible allocation $\pi_f(\gamma)$.

[de] The specific condition is that $f(\gamma) + (7/4 - \gamma)f'(\gamma) \geq 0$ for all $\gamma \in \Gamma$. To establish this point, Amador and Bagwell (2013) refer to the sufficient conditions that they provide for the optimality of interval allocations in the general version of their model.

binding τ^w that Bagwell and Staiger (2005a) characterize indeed describes the optimal trade agreement in the full set of incentive-compatible allocations for this model.[df]

In a second application, Amador and Bagwell (2013) consider a model closely related to the monopolistic competition model described in Section 2.3. Specifically, they consider a short-run version of the model in which the number of home- and foreign-country firms is fixed at the exogenous value $n_h = n_f > 1$, so that firms may enjoy positive profits, and in which the utility functions in (17) are replaced by $U = \log(C_D) + C_Y$ and $U^* = \log(C_D^*) + C_Y$, where as before $C_D = [\sum_i (c_i)^\alpha]^{1/\alpha}$, $C_D^* = [\sum_i (c_i^*)^\alpha]^{1/\alpha}$, $\alpha \in (0, 1)$, and c_i and c_i^* are the respective consumption levels in the home and foreign countries of variety i of the differentiated good. Relative to the partial equilibrium model with perfect competition, a novel feature of this model is that trade is two-way within the differentiated sector, and so the profit earned by each home-country firm depends on the import tariffs in both countries.[dg] To simplify that analysis, Amador and Bagwell assume that the private information is one-sided and concerns the welfare weight that the home-country government attaches to the producer surplus enjoyed by home-country firms.[dh] Despite the more complex trade setting, Amador and Bagwell show that the simple sufficient conditions described earlier can be applied once π is appropriately defined.

In particular, Amador and Bagwell (2013) establish conditions under which the optimal tariff-cap allocation is optimal in the full set of incentive-compatible allocations for the (short-run) monopolistic competition model. The simple sufficient conditions for this model are satisfied if the density is nondecreasing and the parameters are such that $1 > \alpha \geq 2/3$, $\alpha(1-\alpha)\overline{\gamma} < 1$ and $E(\gamma) - \alpha\underline{\gamma} - (2-\alpha)/\alpha > 0$, where $\alpha \geq 2/3$ ensures that $\kappa \geq 1/2$, the role of the latter condition is to generate an interior value for γ_H, and the other parameter restrictions ensure that the model satisfies the basic assumptions of the general delegation problem.[di]

The work described earlier provides a foundation for other recent analyses of tariff caps and binding overhang. Beshkar et al. (2015) extend the linear-quadratic partial-equilibrium

[df] Amador and Bagwell (2013) show also that related results hold in a partial equilibrium model with perfect competition that features log utility and endowments (inelastic supply).

[dg] The dependence takes a separable form, since the import tariff in one country does not affect the terms of trade and nor thereby the price index in the other country. See also Ossa (2012) for a complete-information analysis of trade negotiations in a monopolistic competition model where profits matter due to a fixed number of firms.

[dh] The foreign-country government is assumed to maximize national income. An alternative would be to assume a second differentiated sector and let the foreign government be privately informed about its preferences as well.

[di] Amador and Bagwell (2013) represent home-country government welfare as $W = CS + TR + \gamma PS$, where CS is the consumer surplus enjoyed by home-country consumers, TR is home-country tariff revenue, and PS is the producer surplus enjoyed by home-country firms. They then define π as a specific function of $n_h = n_f$ and the home-country import tariff, τ, where π influences PS but is distinct from PS, and they show that the corresponding formulation of the delegation problem satisfies $v'(\pi) < 0 < v''(\pi)$.

model in an important direction to allow for countries with asymmetric sizes. They focus on tariff caps and provide theoretical and empirical support for the prediction that a higher tariff cap, and thus a greater probability of binding overhang, is optimal when the market power of the importing country is lower. To gain some insight, consider the extreme case of a small country. Since the import tariff for such a country does not generate a terms-of-trade externality, an optimal trade agreement would grant such a country full discretion to respond to its preference shocks. A tariff cap that is sufficiently high accomplishes this objective. Amador and Bagwell (2012) consider the possibility that governments have private information about the value of tariff revenue. They provide an approach under which this problem can be solved using Amador and Bagwell's (2013) propositions for the delegation problem with money burning. For the linear-quadratic model, they then give conditions under which an optimal trade agreement again takes the form of a tariff cap.[dj]

Tariff caps can also be rationalized in other modeling frameworks. Horn et al. (2010), for example, consider a model with contracting costs and show that a weak-binding rule is preferred to a rigid (ie, strong)-binding rule, due to the downward flexibility that the former allows. As in the discussion above, binding overhang occurs with positive probability when a weak-binding rule is used. Maggi and Rodriguez-Clare's (2007) work is also related. They consider a model in which a trade agreement addresses both terms-of-trade and a domestic-commitment problems. While binding overhang does not occur in equilibrium in their model, the potential to apply a tariff below the bound level gives rise to ex post lobbying and thereby helps to diminish the extent of excess investment that occurs in the ex ante stage.[dk]

We also highlight a new literature on the trade effects of tariff bindings, as distinct from reductions in applied tariffs, for settings characterized by policy uncertainty.[dl] As Handley (2014) and Handley and Limão (2015) argue, exporters may be sensitive to policy uncertainty, and tariff caps may thus stimulate entry into export markets by reducing the risk of future increases in protection. Their empirical analyses indicate large trade effects of tariff bindings in trade agreements, both in the context of WTO bindings (for Australia, in Handley, 2014) and preferential-agreement bindings (for Portugal joining the EC, in Handley and Limão, 2015). We view the study of tariff caps in the presence of policy uncertainty as a promising direction for future research.

[dj] In other related work, Bagwell (2009) considers the linear-quadratic model but assumes that the political pressure variable, γ, can only take two values. The optimal trade agreement in the two-type setting does not take the form of a tariff cap. Frankel (forthcoming) is also related. He considers delegation problems with multiple decisions, where each decision has its own underlying state variable, and describes a constant-bias setting where a cap (defined as a ceiling on the weighted average of actions) is optimal. A key theme is that linking decisions can soften incentive constraints. See Athey and Bagwell (2001) and Jackson and Sonnenschein (2007) for broadly related themes in different contexts.

[dk] Another approach is taken by Bowen (2015). Working with a dynamic legislative bargaining model, she considers how tariff caps affect the domestic legislative process and thereby the resulting applied tariffs.

[dl] Other recent contributions to this literature include Limão and Maggi (2015) and Pierce and Schott (2015).

6. CONCLUSION

What does economics have to say about the design of international trade agreements? In this chapter, we have reviewed a literature on this question, providing detailed coverage on three key design features of the GATT/WTO: reciprocity, nondiscrimination as embodied in the MFN principle, and tariff bindings and binding "overhang." Each of these features is central to the design of the GATT/WTO, and we have argued that an economic perspective can go a long way toward revealing a consistent logic to the inclusion of these design features in trade agreements.

We conclude by briefly mentioning several other topics relating to trade-agreement design that are examined in detail elsewhere in this Handbook. A broad theme of our chapter is that the design features of trade agreements are sufficiently deliberate and connected to economic tradeoffs that they can be usefully analyzed from an economic perspective. A good further illustration of this theme can be found in early discussions of the role of dispute settlement procedures to be included in GATT. GATT's dispute settlement procedures raise remarkably subtle tradeoffs between completion of the contract, compliance and compensation that were at the heart of the discussions surrounding these procedures at the time of their initial design.[dm] Themes developed in this chapter provide a foundation for understanding economic aspects of enforcement and dispute settlement procedures in trade agreements, and for understanding as well the manner in which trade agreements can provide upward flexibility to shocks. The themes raised in this chapter also suggest that a properly designed trade agreement must ensure that nontariff and behind-border measures are not used opportunistically, so as to undermine the value of negotiated tariff commitments. Similarly, in multicountry settings, the work described above suggests that discriminatory liberalization may give rise to third-party externalities, a consideration that informs the debate over preferential vs multilateral liberalization. More broadly, other design issues include the treatment of investment and intellectual property in trade agreements; indeed, it is natural to ask what criteria may be put forth to determine which policies are "linked" through trade agreements. These and other design issues are treated in later chapters of this Handbook. These chapters, as they relate to design issues, reinforce the main message of our chapter: economic arguments provide valuable tools for evaluating and interpreting the design of trade agreements.

[dm] For example, in his proposal for a commercial union which would lay much of the groundwork for the design of GATT, James Meade emphasized the tradeoff between writing a more detailed and precise contract vs relying on the dispute settlement system to interpret the contract when disputes inevitably arise (see Meade, 1942, as reproduced in Culbert, 1987, p. 404). And in an early analysis of the 1950–51 GATT Torquay Round, it was suggested that measured compensatory adjustments to restore reciprocity might require on-equilibrium-path authorization of retaliation by GATT at the same time that the possibility of further retaliation and an all-out trade war was to be held as an off-equilibrium threat (see US International Chamber of Commerce, 1955, pp. 63–64).

ACKNOWLEDGMENTS

We thank Emily Blanchard and Giovanni Maggi as our discussants and for detailed comments on an earlier draft and participants at the Handbook of Commercial Policy Conference held at Dartmouth College June 4–June 6 2015 for helpful discussion. K. Bagwell thanks the Center for Advanced Studies in the Behavioral Sciences for support and hospitality.

REFERENCES

Admati, A.R., Perry, M., 1987. Strategic delay in bargaining. Rev. Econ. Stud. 54, 345–364.
Amador, M., Bagwell, K., 2012. Tariff revenue and tariff caps. Am. Econ. Rev. Pap. Proc. 102 (3), 459–465.
Amador, M., Bagwell, K., 2013. The theory of optimal delegation with an application to tariff caps. Econometrica 81 (4), 1541–1599.
Antràs, P., Staiger, R.W., 2012a. Offshoring and the role of trade agreements. Am. Econ. Rev. 102 (7), 3140–3183.
Antràs, P., Staiger, R.W., 2012b. Trade agreements and the nature of price determination. Am. Econ. Rev. Pap. Proc. 102 (3), 470–476.
Athey, S., Bagwell, K., 2001. Optimal collusion with private information. Rand J. Econ. 32 (3), 428–465.
Bacchetta, M., Piermartini, R., 2011. The Value of Bindings. World Trade Organization, Geneva.
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, Huntington, NY, pp. 733–770 (Chapter 40).
Bagwell, K., 2009. Self-enforcing trade agreements and private information. NBER Working Paper No. 14812.
Bagwell K., Bown C., Staiger R.W., forthcoming. Is the wto passé? *J. Econ. Lit.*
Bagwell, K., Lee, S.H., 2015. Trade Policy Under Monopolistic Competition with Firm Selection (Mimeogr.). Stanford University.
Bagwell, K., Mavroidis, P.C., Staiger, R.W., 2007. Auctioning countermeasures in the WTO. J. Int. Econ. 73 (2), 309–332.
Bagwell, K., Staiger, R.W., 1999. An economic theory of GATT. Am. Econ. Rev. 89 (1), 215–248.
Bagwell, K., Staiger, R.W., 2001a. Domestic policies, national sovereignty, and international economic institutions. Q. J. Econ. 116 (2), 519–562.
Bagwell, K., Staiger, R.W., 2001b. Reciprocity, non-discrimination and preferential agreements in the multilateral trading system. Eur. J. Polit. Econ. 17 (2), 281–325.
Bagwell, K., Staiger, R.W., 2001c. Strategic trade, competitive industries and agricultural trade disputes. Econ. Polit. 13 (2), 113–128.
Bagwell, K., Staiger, R.W., 2002. The Economics of the World Trading System. The MIT Press, Cambridge, MA.
Bagwell, K., Staiger, R.W., 2005a. Enforcement, private political pressure and the GATT/WTO escape clause. J. Leg. Stud. 34 (2), 471–513.
Bagwell, K., Staiger, R.W., 2005b. Multilateral trade negotiations, bilateral opportunism and the rules of GATT/WTO. J. Int. Econ. 67 (2), 268–294.
Bagwell, K., Staiger, R.W., 2006. Will international rules on subsidies disrupt the world trading system? Am. Econ. Rev. 96 (3), 877–895.
Bagwell, K., Staiger, R.W., 2010a. Backward stealing and forward manipulation in the WTO. J. Int. Econ. 82 (1), 49–62.
Bagwell, K., Staiger, R.W., 2010b. The WTO: theory and practice. Annu. Rev. Econ. 2, 223–256.
Bagwell, K., Staiger, R.W., 2011. What do trade negotiators negotiate about? Empirical evidence from the World Trade Organization. Am. Econ. Rev. 101 (4), 1238–1273.
Bagwell, K., Staiger, R.W., 2012a. The economics of trade agreements in the linear cournot delocation model. J. Int. Econ. 88 (1), 32–46.

Bagwell, K., Staiger, R.W., 2012b. Profit shifting and trade agreements in imperfectly competitive markets. Int. Econ. Rev. 53 (4), 1067–1104.
Bagwell, K., Staiger, R.W., 2014. Can the Doha Round be a development round? Setting a place at the table. In: Feenstra, R.C., Taylor, A.M. (Eds.), Globalization in an Age of Crisis: Multilateral Economic Cooperation in the Twenty-First Century. University of Chicago Press for the NBER, Chicago, IL, pp. 91–124.
Bagwell, K., Staiger, R.W., 2015. Delocation and trade agreements in imperfectly competitive markets. Res. Econ. 69 (2), 132–156.
Bagwell, K., Staiger, R.W., Yurukoglu, A., 2015. Multilateral Trade Bargaining: A First Look at the GATT Bargaining Records (Mimeogr.). Stanford University
Baldwin, R., 1987. Politically realistic objective functions and trade policy PROFs and tariffs. Econ. Lett. 24 (3), 287–290.
Baldwin, R.E., Clarke, R.N., 1987. Game-modeling multilateral trade negotiations. J. Policy Model 9 (2), 257–284.
Beckett, G., 1941. The Reciprocal Trade Agreements Program. Columbia University Press, New York, NY.
Beshkar, M., 2010. Trade skirmishes and safeguards: a theory of the WTO dispute settlement process. J. Int. Econ. 82 (1), 35–48.
Beshkar, M., Bond, E.W., Rho, Y., 2015. Tariff binding and overhang: theory and evidence. J. Int. Econ. 97 (1), 1–13.
Binmore, K., 2014. Bargaining and fairness. Proc. Natl. Acad. Sci. USA 111 (Suppl. 3), 10785–10788.
Blanchard, E.J., 2007a. Foreign direct investment, endogenous tariffs, and preferential trade agreements. BE J. Econ. Anal. Policy 7 (1)(Advances), Art. no. 54.
Blanchard, E.J., 2007b. Reevaluating the Role of Trade Agreements: Does Investment Globalization Make the WTO Obsolete? (Mimeogr.). University of Virginia.
Blanchard, E.J., 2010. Reevaluating the role of trade agreements: does investment globalization make the WTO obsolete? J. Int. Econ. 82 (1), 63–72.
Blanchard, E.J., Matschke, X., 2015. US multinationals and preferential market access. Rev. Econ. Stat. 97 (4), 839–854.
Bowen, T.R., 2015. Legislated protection and the World Trade Organization. Int. Econ. Rev. 56 (4), 1349–1384.
Brainard, L., 1994. Last one out wins: trade policy in an international exit game. Int. Econ. Rev. 35 (1), 151–172.
Brander, J.A., 1995. Strategic trade policy. In: Grossman, G.M., Rogoff, K. (Eds.), Handbook of International Economics, 3. North-Holland, Amsterdam, NL, pp. 1395–1455.
Brock, W.A., Magee, S.P., 1978. The economics of special interest politics. Am. Econ. Rev. 68 (2), 246–250.
Brou, D., Ruta, M., 2013. A commitment theory of subsidy agreements. BE J. Econ. Anal. Policy 13 (1), 239–270.
Campolmi, A., Fadinger, H., Forlati, C., 2014. Trade policy: home market effect versus terms-of-trade externality. J. Int. Econ. 93, 92–107.
Carmichael, C.M., 1987. The control of export credit subsidies and its welfare consequences. J. Int. Econ. 23 (1-2), 1–19.
Caves, R.E., 1976. Economic models of political choice: Canada's tariff structure. Can. J. Econ. 9 (2), 278–300.
Choi, J.P., 1995. Optimal tariffs and the choice of technology: discriminatory tariffs vs. the most favored nation clause. J. Int. Econ. 38 (1-2), 143–160.
Costinot, A., Rodriguez-Clare, A., Werning, I., 2015. Micro to Macro: Optimal Trade Policy with Firm Heterogeneity (Mimeogr.). MIT.
Cramton, P., 1992. Strategic delay in bargaining with two-sided uncertainty. Rev. Econ. Stud. 59 (1), 205–225.
Culbert, J., 1987. War-time anglo-american talks and the making of the GATT. World Econ. 10 (4), 381–408.
DeRemer, D.R., 2012. Essays on International Trade Agreements Under Monopolistic Competition (Dissertation). Columbia University.

DeRemer, D.R., 2013. Domestic Policy Coordination in Imperfectly Competitive Markets (Mimeogr.). Hungarian Academy of Sciences.
Dhingra, S., 2014. Reconciling observed tariffs and the median voter model. Econ. Polit. 26 (3), 483–504.
Dixit, A., 1987. Strategic aspects of trade policy. In: Bewley, T.F. (Ed.), Advances in Economic Theory: Fifth World Congress. Cambridge University Press, New York, NY, pp. 329–362.
Dixit, A., Stiglitz, J.E., 1977. Monopolistic competition and optimum product diversity. Am. Econ. Rev. 67 (3), 297–308.
Dutt, P., Mitra, D., 2002. Endogenous trade policy through majority voting: an empirical investigation. J. Int. Econ. 58 (1), 107–133.
Ederington, J., 2001. International coordination of trade and domestic policies. Am. Econ. Rev. 91 (5), 1580–1593.
Feenstra, R.C., Bhagwati, J.N., 1982. Tariff seeking and the efficient tariff. In: Bhagwati, J.N. (Ed.), Import Competition and Response. University of Chicago Press for the NBER, Chicago, IL, pp. 245–262.
Findlay, R., Wellisz, S., 1982. Endogenous tariffs, the political economy of trade restrictions, and welfare. In: Bhagwati, J.N. (Ed.), Import Competition and Response. University of Chicago Press for the NBER, Chicago, IL, pp. 223–244.
Finger, J.M., 1979. Trade liberalization: a public choice perspective. In: Amacher, R., Haberler, G., Willett, T. (Eds.), Challenges to a Liberal International Economic Order. American Enterprise Institute, Washington, DC, pp. 421–453.
Frankel A., forthcoming. Delegating multiple decisions. Am. Econ. J. Microecon.
GATT, 1958. Trends in International Trade: A Report by a Panel of Experts. GATT, Geneva.
Grossman, G.M., 2016. The purpose of trade agreements. In: Bagwell, K., Staiger, R.W. (Eds.), The Handbook of Commercial Policy, vol. 1A. Elsevier, Netherlands, Chapter 7, pp. 379–434.
Grossman, G.M., Helpman, E., 1994. Protection for sale. Am. Econ. Rev. 84 (4), 833–850.
Grossman, G.M., Helpman, E., 1995. Trade wars and trade talks. J. Polit. Econ. 103 (4), 675–708.
Grossman, G.M., Sykes, A.O., 2010. Optimal retaliation in the WTO–a commentary on the upland cotton arbitration. World Trade Rev. 10, 133–164.
Gulotty, R., 2014. Governing Trade Beyond Tariffs: The Politics of Multinational Production and Its Implications for International Cooperation (Dissertation). Stanford University.
Hagerty, K.M., Rogerson, W.P., 1987. Robust trading mechanisms. J. Econ. Theory 42, 94–107.
Handley, K., 2014. Exporting under trade policy uncertainty: theory and evidence. J. Int. Econ. 94 (1), 50–66.
Handley, K., Limão, N., 2015. Trade and investment under policy uncertainty: theory and firm evidence. Am. Econ. J. Econ. Policy 7 (4), 189–222.
Harstad, B., 2007. Harmonization and side payments in political cooperation. Am. Econ. Rev. 97 (3), 871–889.
Helpman, E., Krugman, P.E., 1989. Trade Policy and Market Structure. The MIT Press, Cambridge, MA.
Hillman, A.L., 1982. Declining industries and political-support protectionist motives. Am. Econ. Rev. 72 (5), 1180–1187.
Hoda, A., 2001. Tariff Negotiations and Renegotiations under the GATT and the WTO: Procedures and Practices. WTO and Cambridge University Press, New York.
Hoekman, B., Kostecki, M., 1995. The Political Economy of the World Trading System. Oxford University Press, Oxford.
Holmstrom, B., 1977. On Incentives and Control in Organizations (Ph.D. dissertation). Stanford University.
Horn, H., Maggi, G., Staiger, R.W., 2010. Trade agreements as endogenously incomplete contracts. Am. Econ. Rev. 100 (1), 394–419.
Interim Commission for the ITO, 1949. The Attack on Trade Barriers: A Progress Report on the Operation of the General Agreement on Tariffs and Trade, Geneva, August.
Jackson, J.H., 1969. World Trade and the Law of GATT. Bobbs-Merrill, New York, NY.
Jackson, M.O., Sonnenschein, H.F., 2007. Overcoming incentive constraints by linking decisions. Econometrica 75 (1), 241–257.
Jawara, F., Kwa, A., 2003. Behind the Scenes at the WTO: The Real World of International Trade Negotiations. Zed Books, London.
Johnson, H.G., 1953-1954. Optimum tariffs and retaliation. Rev. Econ. Stud. 21 (2), 142–153.

Karacaovali, B., Limão, N., 2008. The clash of liberalizations: preferential vs. multilateral trade liberalization in the European Union. J. Int. Econ. 74 (2), 299–327.
Kennan, J., Riezman, R., 1988. Do big countries win tariff wars? Int. Econ. Rev. 29 (1), 81–85.
Krugman, P.R., 1997. What should trade negotiators negotiate about? J. Econ. Lit. 35 (1), 113–120.
Lapan, H.E., 1988. The optimal tariff, production lags, and time consistency. Am. Econ. Rev. 78 (3), 395–401.
Lawrence, R.Z., 2003. Crimes and Punishments? Retaliation Under the WTO. Institute for International Economics, Washington, DC.
Lee, G.M., 2007. Trade agreements with domestic policies as disguised protection. J. Int. Econ. 71 (1), 241–259.
Lee, G.M., 2014. Optimal International Agreement and Restriction on Domestic Efficiency (Mimeogr.). Singapore Management University.
Limão, N., 2005. Trade policy, cross-border externalities and lobbies: do linked agreements enforce more cooperative outcomes? J. Int. Econ. 67 (1), 175–199.
Limão, N., 2006. Preferential trade agreements as stumbling blocks for multilateral trade liberalization: evidence for the U.S. Am. Econ. Rev. 96 (3), 896–914.
Limão, N., 2007. Are preferential trade agreements with non-trade objectives a stumbling block for multilateral liberalization? Rev. Econ. Stud. 74 (3), 821–855.
Limão, N., Maggi, G., 2015. Uncertainty and trade agreements. Am. Econ. J.: Microecon. 7 (4), 1–42.
Limão, N., Saggi, K., 2008. Tariff retaliation versus financial compensation in the enforcement of international trade agreements. J. Int. Econ. 76 (1), 48–60.
Ludema, R.D., 1991. International trade bargaining and the most-favored nation clause. Econ. Polit. 3 (1), 1–20.
Ludema, R.D., Mayda, A.M., 2013. Do terms-of trade effects matter for trade agreements? Theory and evidence from WTO countries. Q. J. Econ. 128 (4), 1837–1893.
Ludema, R.D., Zhi, Y., 2015. Tariff Pass-through, Firm Heterogeneity and Product Quality (Mimeogr.). Georgetown University.
Maggi, G., 1999. The role of multilateral institutions in international trade cooperation. Am. Econ. Rev. 89 (1), 190–214.
Maggi, G., 2014. International trade agreements. In: Gopinath, G., Helpman, E., Rogoff, K. (Eds.), The Handbook of International Economics. vol. 4. Elsevier, Amsterdam.
Maggi, G., Rodriguez-Clare, A., 1998. The value of trade agreements in the presence of political pressures. J. Polit. Econ. 106, 574–601.
Maggi, G., Rodriguez-Clare, A., 2007. A political-economy theory of trade agreements. Am. Econ. Rev. 97 (4), 1374–1406.
Maggi, G., Staiger, R.W., 2015a. Optimal design of trade agreements in the presence of renegotiation. Am. Econ. J. Microecon. 7 (1), 109–143.
Maggi, G., Staiger, R.W., 2015b. Trade Disputes and Settlement (Mimeogr.). Yale University.
Matsuyama, K., 1990. Perfect equilibria in a trade liberalization game. Am. Econ. Rev. 80 (3), 480–492.
Mavroidis, P.C., 2016. Dispute settlement in the WTO: mind over matter. In: Bagwell, K., Staiger, R.W. (Eds.), The Handbook of Commercial Policy, vol. 1A. Elsevier, Netherlands, Chapter 6, pp. 333–377.
Mayer, W., 1981. Theoretical considerations on negotiated tariff adjustments. Oxf. Econ. Pap. 33 (1), 135–153.
Mayer, W., 1984. Endogenous tariff formation. Am. Econ. Rev. 74 (5), 970–985.
McCalman, P., 2002. Multi-lateral trade negotiations and the most-favored nation clause. J. Int. Econ. 57 (1), 151–176.
Melitz, M., Ottaviano, G., 2008. Market size, trade, and productivity. Rev. Econ. Stud. 75 (1), 295–316.
Mill, J., 1844. Essays on Some Unsettled Questions of Political Economy. Parker, London.
Myerson, R.B., Satterthwaite, M.A., 1983. Efficient mechanisms for bilateral trading. J. Econ. Theory 29, 265–281.
Olson, M., 1965. The Logic of Collective Action: Public Goods and the Theory of Groups. Harvard University Press, Cambridge, MA.
Ornelas, E., 2016. Special and differential treatment for developing countries. In: Bagwell, K., Staiger, R.W. (Eds.), The Handbook of Commercial Policy, vol. 1B. Elsevier, Netherlands, Chapter 7, pp. 369–432.

Ossa, R., 2011. A 'new-trade' theory of GATT/WTO negotiations. J. Polit. Econ. 119 (1), 122–152.
Ossa, R., 2012. Profits in the new trade approach to trade negotiations. Am. Econ. Rev. Pap. Proc. 102 (2), 466–469.
Ossa, R., 2014. Trade wars and trade talks with data. Am. Econ. Rev. 104 (12), 4104–4146.
Penrose, E., 1953. Economic Planning for the Peace. Princeton University Press, Princeton, NJ.
Pierce, J., Schott, P.K., 2015. The Surprisingly Swift Decline of U.S. Manufacturing Employment (Mimeogr.). Yale University.
Potipiti, T., 2012. Import Tariffs and Export Subsidies in the WTO: A Small Country Approach (Mimeogr.). Chulalongkorn University.
Saggi, K., 2004. Tariffs and the most favored nation clause. J. Int. Econ. 63 (2), 341–368.
Schwartz, W.F., Sykes, A.O., 1997. The economics of the most favored nation clause. In: Bhandari, J.S., Sykes, A.O. (Eds.), Economic Dimensions in International Law: Comparative and Empirical Perspectives. Cambridge University Press, Cambridge, UK, pp. 43–79.
Shirono, K., 2004. Are WTO Tariff Negotiations Reciprocal? An Analysis of Tariff Liberalization (Mimeogr.). Columbia University.
Spagnola, G., 1999a. Issue Linkage, Delegation and International Policy Coordination (Mimeogr.). Stockholm School of International Economics.
Spagnola, G., 1999b. On interdependent supergames: multimarket contact, concavity and collusion. J. Econ. Theory 89 (1), 127–139.
Staiger, R.W., 2006. What can developing countries achieve in the WTO? J. Econ. Lit. 44, 779–795.
Staiger, R.W., Sykes, A.O., 2011. International trade, national treatment, and domestic regulation. J. Leg. Stud. 40 (1), 149–203.
Staiger, R.W., Sykes, A.O., forthcoming. How important can the non-violation clause before the GATT/WTO? Am. Econ. J. Microecon.
Staiger, R.W., Tabellini, G., 1987. Discretionary trade policy and excessive protection. Am. Econ. Rev. 77 (5), 823–837.
Staiger, R.W., Tabellini, G., 1989. Rules and discretion in trade policy. Eur. Econ. Rev. 33 (6), 1265–1277.
Suwanprasert, W., 2014. The Role of MFN in the 'New Trade' Model (Mimeogr.). University of Wisconsin.
Sykes, A.O., 2005. The economics of WTO rules on subsidies and countervailing measures. In: Appleton, A. E., Macrory, P.F.J., Plummer, M.G. (Eds.), The World Trade Organization: Legal, Economic and Political Analysis, Springer-Verlag, USA.
Tornell, A., 1991. On the ineffectiveness of made-to-measure protectionist programs. In: Helpman, E., Razin, A. (Eds.), International Trade and Trade Policy. MIT Press, Cambridge, pp. 66–79.
Torrens, R., 1844. The Budget: On Commercial Policy and Colonial Policy. Smith, Elder, London.
United States Council of the International Chamber of Commerce, 1955. G.A.T.T.: an analysis and appraisal of the General Agreement on Tariffs and Trade. Prepared by the William L. Clayton Center for International Economic Affairs of the Fletcher School of Law and Diplomacy, February.
Venables, A.J., 1985. Trade and trade policy with imperfect competition: the case of identical products and free entry. J. Int. Econ. 19 (1-2), 1–19.
Venables, A.J., 1987. Trade and trade policy with differentiated products: a Chamberlinian-Ricardian model. Econ. J. 97 (387), 700–717.
WTO, 2009. World Trade Report 2009: Trade Policy Commitments and Contingency Measures. WTO, Geneva.
WTO Arbitrators, 1999. European communities–measures concerning meat and meat products (hormones); original complaint by Canada; recourse to arbitration by the European communities under article 22.6 of the DSU: decision by the arbitrators. WT/DS48/ARB, 12 July.