Backward stealing and forward manipulation in the WTO

Kyle Bagwell, Robert W. Staiger *

Department of Economics, Stanford University, Stanford CA, 94305 United States; and NBER

A R T I C L E   I N F O

Article history:
Received 27 March 2007
Received in revised form 12 April 2010
Accepted 14 April 2010

JEL classification:
F13

Keywords:
GATT
WTO
Trade negotiations

A B S T R A C T

Motivated by the structure of WTO negotiations, we analyze a bargaining environment in which negotiations proceed bilaterally and sequentially under the most-favored-nation (MFN) principle. We identify backward-stealing and forward-manipulation problems that arise when governments bargain under the MFN principle in a sequential fashion. We show that these problems impede governments from achieving the multilateral efficiency frontier unless further rules of negotiation are imposed. We identify the WTO nullification-or-impairment and renegation provisions and its reciprocity norm as rules that are capable of providing solutions to these problems. In this way, we suggest that WTO rules can facilitate the negotiation of efficient multilateral trade agreements in a world in which the addition of new and economically significant countries to the world trading system is an ongoing process.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Under the auspices of the World Trade Organization (WTO) – and GATT, its predecessor organization created in 1947 – governments have met with broad success in liberalizing world trade. This success is especially remarkable in light of three prominent features of the GATT/WTO negotiating environment. First, WTO negotiations abide by the most-favored-nation (MFN) principle, under which a WTO-member country must provide all member-countries with the same conditions of access to its markets. Second, WTO negotiations take place overwhelmingly among small numbers of countries. And third, GATT/WTO negotiations have extended over more than half a century, during which time the addition of new countries to the world trading system – via either the process of economic development or the act of accession to the GATT/WTO – has occurred on a continuing basis. As a consequence of these three features, it is routine for a country to engage in market access negotiations on a product with one country, having previously negotiated tariff commitments on that product with another country, all subject to MFN.

In this sequential MFN negotiating environment, a pair of potential impediments to multilateral efficiency may be identified. First, under MFN, any market access concession that a country makes to an early negotiating partner is automatically available to future negotiating partners as well. To reduce the associated potential for free-riding, a country might then engage in inefficient “foot-dragging,” offering little to early negotiating partners in order to maintain its bargaining position for later negotiations. A second impediment to efficiency might arise if later negotiating partners themselves engage in “bilateral opportunism” and seek to alter the market access implications of earlier negotiations to their own advantage. More broadly, we may associate the first impediment with forward manipulation, in which early agreements are manipulated to alter the outcome of later negotiations, and the second impediment with backward stealing, in which early agreements are structured to take surplus from earlier negotiating partners.

Does the GATT/WTO owe its success to the fact that these potential impediments are simply unimportant? Or can its rules instead be credited with providing governments with some assurance that forward-manipulation and backward-stealing problems will not become severe? In this paper, we suggest that the potential impediments to efficiency created by these problems are important. And we identify GATT/WTO rules that can help governments overcome these impediments.

Our analysis is carried out within a three-country two-good world, in which a home-country government negotiates bilaterally and sequentially with each of two trading partners, subject to the MFN principle.
We also permit governments to make direct international transfers as part of their bilateral negotiations. We do this for three reasons. The first reason is to ensure analytical tractability: the feasibility of direct international transfers simplifies our analysis considerably. The second reason is pedagogical: in the presence of international transfers, it is clear that static bargaining (among all three countries) would yield efficient outcomes, and this permits us to highlight the issue of dynamic bargaining inefficiencies that is the focus of our analysis. And the third reason is to endow governments with a reasonably flexible portfolio of policy instruments. While it is not standard for GATT/WTO trade negotiations to involve explicit transfers as part of the agreement, these negotiations do often involve more than just tariff reductions. Our assumption that direct international transfers are feasible may be seen as an attempt to capture these additional policy dimensions in a simple model, with “reality” positioned somewhere in between the extremes of negotiations over tariffs only and negotiations over tariffs and direct international transfers.

Within this framework, we characterize the multilateral efficiency frontier, and we then explore whether this frontier can be reached in subgame perfect equilibria of specific bargaining games that entail sequential and bilateral negotiations under MFN. We explore this issue in two broad steps. We first show that, in our basic sequential MFN bargaining game, the backward-stealing problem makes it impossible for governments to reach the multilateral efficiency frontier: beginning from any efficient combination of tariffs and transfers, the home government and its later negotiating partner can always alter the tariffs and transfers under their control in a way that benefits them at the expense of the (unrepresented) early negotiating partner. When we impose an exogenous “security requirement” that later agreements may not involve backward stealing, we find that the forward-manipulation problem makes it generally impossible for governments to reach the efficiency frontier: as a general matter, the home government can engage in inefficient foot-dragging with its early negotiating partner by keeping its tariff high, and both the home government and its early negotiating partner can thereby benefit at the expense of the (unrepresented) later negotiating partner, who is stuck with a less-favorable disagreement point.

With the backward-stealing and forward-manipulation problems identified in our basic sequential MFN bargaining game, we then turn to the second broad step of our analysis. We demonstrate that renegotiation opportunities such as those provided in the GATT/WTO can curtail the significance of early negotiation outcomes for the disagreement payoffs of subsequent negotiating partners, and thereby alleviate the inefficiency associated with forward manipulation. And we show that the GATT/WTO reciprocity norm and nullification-or-impairment provisions can mimic a security requirement, and thereby can be seen as helping to alleviate the backward-stealing problem. When joined together, we establish that these rules permit governments to achieve the multilateral efficiency frontier in a wide range of circumstances.

Our main finding is then that the GATT/WTO rules analyzed in this second step permit governments engaged in sequential MFN bargaining to achieve efficient outcomes that are otherwise precluded by the backward-stealing and forward-manipulation problems identified in step one. But this finding raises a further question: Is there any evidence that these GATT/WTO rules were adopted at least in part for the purpose of addressing the problems that our analysis suggests they are well-designed to solve?

In this regard, it is widely acknowledged (e.g., Rhodes, 1993) that much of the GATT architecture was inspired by United States experience under the 1934 Reciprocal Trade Agreements Act (RTAA). What is less well-appreciated is the way in which the RTAA was itself influenced by the successes and failures of the many international attempts that came before it to address the problem of high trade barriers. In fact, as we describe in more detail elsewhere (see Bagwell and Staiger, 2010a), at a broad level by 1934 the United States was (a) frustrated with its experience in multilateral tariff bargaining, and (b) based on European experience, well-aware that the effectiveness of bilateral tariff bargaining could be undone by the twin problems of “concession erosion” (backward stealing) and “bargaining tariffs” (forward manipulation). And so, under the RTAA, the United States sought to devise a method of bilateral tariff bargaining that would not be susceptible to these fundamental problems. The central elements of this method included unconditional MFN as well as tactics for preserving reciprocity in bilateral negotiations; and as we have observed, the GATT/WTO adopted the essential approach embodied in the RTAA.

While our paper is broadly related to a literature in Industrial Organization on contracting with externalities (see, for example, Aghion and Bolton, 1987, McAfee and Schwartz, 1994, and Marx and Shaffer, 2004, 2008), in the International Trade literature there are two papers that are most closely related to the present analysis. The first paper is Bagwell and Staiger (2005). In that paper, we are also concerned with the possibility of inefficient negotiating outcomes when pairs of countries can negotiate bilaterally. But there are two important differences between that paper and the present analysis. First, in our earlier paper we identify rules of negotiation that serve to protect the welfare of governments that are not participating in a bilateral negotiation, and we relate these rules to WTO principles, but we do not ask the central question of the present analysis: starting from an inefficient (non-cooperative) set of policies, can a simple set of rules be identified which (i) allow governments who engage in sequential bilateral MFN negotiations to arrive at an efficient arrangement, and (ii) have a counterpart in GATT articles? Providing an answer to this question requires a model of the sequential bargaining process, something that our earlier paper does not provide. A second important difference is that we do not permit direct international transfers in our earlier paper. We indicate below how the possibility of international transfers affects our earlier results. The second related paper is Limao (2007), who explores an idea related to our foot-dragging result. He shows that a government may engage in foot-dragging under MFN to enhance its bargaining position with regard to a subsequent negotiating partner. However, in Limao’s model, foot-dragging arises in anticipation of a subsequent preferential agreement with non-trade objectives, while in our model foot-dragging arises in anticipation of subsequent MFN market access negotiations.

The next section introduces the three-country two-good model and characterizes the efficiency frontier. Section 3 introduces the basic sequential MFN bargaining game, and identifies the backward-stealing problem, while Section 4 identifies the forward-manipulation problem. Sections 5 and 6 introduce renegotiation opportunities and nullification-or-impairment/reciprocity provisions as a means by which to alleviate the forward-manipulation and backward-stealing problems, respectively. Section 7 concludes.

2. The model

2.1. The basic setup

We consider a perfectly competitive general equilibrium environment. We assume that country $A$ exports goods to countries $B$ and $C$.

---

4. For example, the Agreement on Trade-Related Aspects of Intellectual Property Rights negotiated in the Uruguay Round is often interpreted as a transfer from the developing world to industrialized countries that was granted in exchange for certain market access concessions (such as the phase-out of the Multifiber Arrangement).

5. As will become clear below, it is important that the exporting government’s negotiated trade policy commitments can stimulate its exports. In a general equilibrium setting this feature follows from Lerner symmetry even when negotiations only cover import tariffs (for recent empirical confirmation that import tariffs act as significant export taxes, see for example Irwin, 2007, and Tokarick, 2007). In a partial equilibrium setting this feature would require that the exporting government negotiate directly over export policies, which runs counter to GATT/WTO practice. Hence, while working in a partial equilibrium model would simplify our analysis, we prefer to adopt a general equilibrium setting so that we may maintain our focus on negotiations over import tariffs.
in exchange for imports of good $x$ from B and C. Country A may levy an MFN import tariff $\tau^A$, while countries B and C may each levy their own import tariff, $\tau^B$ and $\tau^C$, respectively. We adopt the convention that $\tau^j$ represents one plus the ad valorem import tariff of country $j$, and we let $\tau_j$ denote the vector of tariffs $(\tau^A, \tau^B, \tau^C)$. Country A may also make direct (consumption) transfers to country B and/or country C. We denote the (positive or negative) transfer from A to B by $\hat{\tau}_j^A$ and from A to C by $\hat{\tau}_j^C$, measured in units of good $y$. The total net transfers made from A to its trading partners is then $\hat{\tau}_j^A = \hat{\tau}_j^B + \hat{\tau}_j^C$, and we let $\tau$ denote the vector of transfers $(\tau^A, \tau^B, \tau^C)$.

Provided that country A's (MFN) tariff does not prohibit trade with either of its trading partners B and C, there will be a common exporter of good $x$ in countries B and C, and we denote this price by $P_{wE}^j$. The export price for good $y$ in country A is denoted by $P_C^A$. We may define the ratio of “world” prices (relative exporter prices) as $P_{wE}^j = P_C^j$. We refer to $P_{wE}^A$ as the world price or the terms of trade between country A and its trading partners B and C. Similarly, we let $P_{wE}^B$ denote the price of good $y$ relative to the price of good $y$ prevailing locally in country $j \in \{A, B, C\}$. We refer to $P_{wE}^j$ as the ratio of local prices in country $j$.

With non-proportional transfers, international arbitrage links world and local prices: $P^j = \tau_j^A P_{wE}^j \equiv \hat{\tau}_j^A$; $P^j = \tau_j^B P_{wE}^j \equiv \hat{\tau}_j^B$; $P^j = \tau_j^C P_{wE}^j \equiv \hat{\tau}_j^C$ for $j \in \{B, C\}$.

We assume that the international transfers have no secondary burden or blessing (i.e., that they do not affect the equilibrium terms of trade). In each country, the sum of net transfers and tariff revenue is distributed to consumers in a lump-sum fashion.

For any world price, each country’s trade must balance in light of its net transfers: $P_{wE}^j M_j^A = E^A - \hat{\tau}_j^A$ and $M_j - \hat{\tau}_j = P_{wE}^j E^j$ for $j \in \{B, C\}$.

The $y$-market is assumed to clear at $P_{wE}^j$ by Eqs. (1)–(2). We assume that the Marshall–Lerner stability conditions are met globally (ensuring a unique $P_{wE}^j$ given $\tau_j$), so that an inward shift of a country’s import demand curve improves its terms of trade, and that the Lerner and Metzler paradoxes are ruled out, so that $\partial P_{wE}^j / \partial x > 0$, $\partial P_{wE}^j / \partial x < 0$, $\partial P_{wE}^j / \partial \hat{\tau}_j > 0$, and $\partial P_{wE}^j / \partial \hat{\tau}_j < 0$ for $j \in \{B, C\}$.

With $P_{wE}^j$ held fixed, a change in $P_{wE}^j$ or $\hat{\tau}_j$ can affect $M_j^A$ only through the effect on A’s national income. But the income effect of a small change in $P_{wE}^j$, measured in units of good $y$, is given by the import volume $M_j^A$. With analogous observations for B and C, we thus record the following structure on each country’s trade function (subscripts denote partial derivatives):

$$M_{wE}^j = \left[\begin{array}{c} M_{wE}^A \times M^A \times \hat{\tau}_j^A \times \hat{\tau}_j^B \times \hat{\tau}_j^C \times E^A \times E^B \times E^C \end{array}\right]$$

Finally, we represent the objectives of each government as a general function of its local prices, its terms of trade, and the net transfers it grants or receives. In particular, we represent the welfare of the government of country $j$ by $\hat{W}_j^A(\tau^A, \hat{\tau}_j^A, P_{wE}^A, E^A)$ for $j \in \{A, B, C\}$. We place the following basic restrictions on these objective functions. First, under analogous reasoning to that which leads to (2a), we impose the following structure on each country’s objective function:

$$\hat{\tau}_j^A = \hat{W}_j^A \times \hat{\tau}_j^A \times M^A; \quad \hat{\tau}_j^B = \hat{W}_j^B \times E^B; \quad \hat{\tau}_j^C = \hat{W}_j^C \times E^C$$

As before, this structure reflects the link between direct international transfers and the income effects of changes in $P_{wE}$. And second, we assume that, holding its local prices and its terms of trade fixed, each government would prefer an increase in net transfers toward it: $\hat{W}_B^A > 0; \hat{W}_C^B > 0; \hat{W}_C^C > 0$. Under (3), this implies as well that, holding its local prices and its net transfer fixed, each country would prefer a terms-of-trade improvement: $\hat{W}_C^B < 0; \hat{W}_B^C < 0; \hat{W}_C^C > 0$. As we have argued extensively elsewhere (see Bagwell and Staiger, 1999), by leaving government preferences over local prices unspecified, our representation of government objectives is very general and is consistent with national-income-maximizing governments as well as governments that are motivated by various political/distributional concerns.

Our three tariffs and two transfers provide one degree of freedom in achieving any level of welfare for the three governments. This means that any welfare triple can be achieved with an arbitrary market-clearing world price or with any one instrument set at an arbitrary level.

As a feature is important later, we record it in the following lemma (proved in the Appendix):

**Lemma 1.** Any welfare triple can be achieved with an arbitrary market-clearing world price or with any one instrument set at an arbitrary level.

### 2.2. The efficiency frontier

Defining $W^i(\tau, \hat{\tau}) = W^i(\tau \hat{\tau}, P_{wE}(\tau, \hat{\tau}), P_{wE}(\tau, \hat{\tau}), E^i)$, we may now characterize the efficiency frontier. We define the efficiency frontier with respect to the governments’ own preferences, and it is characterized by the set of solutions to:

$$\max_{(\tau, \hat{\tau}, t)} \quad W^A(\tau, \hat{\tau}_j^A + t^A)$$

s.t. $W^B(\tau, \hat{\tau}_j^B) \geq \bar{W}_B; \quad W^C(\tau, t^C) \geq \bar{W}_C$

where $\bar{W}_B$ and $\bar{W}_C$ denote the welfare of the governments of countries B and C, respectively, evaluated at the efficient policies. The five first-order conditions that characterize the efficient selection of $(\tau, \hat{\tau}_j^j, \hat{\tau}_j^C)$, given $\bar{W}_B$ and $\bar{W}_C$, can be written as:

$$\frac{W^A_{\tau}}{W^A_{\hat{\tau}_j^A}} - \frac{W^B_{\tau}}{W^B_{\hat{\tau}_j^B}} - \frac{W^C_{\tau}}{W^C_{\hat{\tau}_j^C}} = 0; \quad \frac{W^A_{\hat{\tau}_j^A}}{W^A_{\tau}} = 0; \quad \frac{W^B_{\hat{\tau}_j^B}}{W^B_{\tau}} = 0; \quad \frac{W^C_{\hat{\tau}_j^C}}{W^C_{\tau}} = 0$$

$$W^B(\tau, \hat{\tau}_j^B) = \bar{W}_B; \quad \text{and} \quad W^C(\tau, \hat{\tau}_j^C) = \bar{W}_C.$$
In words, efficiency condition (4) states that, for \( j = A, B, C \) respectively, a small change in \( \tau^j \) which is accompanied by the change in \( t^j \) that keeps \( B \) indifferent and the change in \( t^j \) that keeps \( C \) indifferent must keep \( A \) indifferent as well.

Throughout the paper we restrict our focus to efficient policy combinations that call for tariffs positioned below the reaction curves of each country, and we ask whether such policy combinations can be implemented as equilibria of specific bargaining games. This below-the-reaction-curves restriction comes with little loss of generality. In each of the games we consider -- as in GATT/WTO negotiations -- governments agree to bind their tariffs at specified levels, and these bindings then place upper limits on permissible tariff choices. As a consequence, any efficient combination of policies that required at least one country to set its tariff above its reaction curve would be unattainable in the bargaining games we consider, provided only that subsequent to the conclusion of negotiations each government is permitted (as we assume) to set its tariff unilaterally subject to the constraint that it does not exceed its negotiated tariff binding.

Efficiency might be achieved with a subset of countries on their tariff reaction curves, but the unilateral nature of the tariff commitments that efficiency would require of the remaining countries is at odds with the "reciprocal" nature of GATT/WTO tariff negotiations.9 Rather than make these arguments repeatedly throughout the paper, we focus from the beginning on efficient policy combinations that call for tariffs positioned below the reaction curves of each country. We record this restriction as:

\[
dW^j / d\tau^j > 0, \quad j \in \{A, B, C\} \quad (A1)
\]

In addition to (A1), we restrict our focus as well to efficient points that satisfy:

\[
\text{sign}(dW^B/d\tau^A) = \text{sign}(dW^C/d\tau^A)
\]

At an efficient point satisfying (A2), B and C agree on the direction (if any) that each would like \( \tau^j \) to move. Exploring cases where the incentives of B and C are opposed might also be of interest, but the aligned case is a natural starting point for analyzing MFN tariff bargaining between A and each of its trading partners.

We treat (A1)–(A2) as maintained assumptions throughout the paper that define the relevant region of the efficiency frontier. These assumptions imply the direction in which each government would prefer each policy to move beginning from an efficient point. In the Appendix we prove:

**Lemma 2.** At any efficient point, the following restrictions apply:

1. \( dW^j / d\tau^j > 0, \quad j \in \{A, B, C\} \);
2. \( dW^j / d\tau^A = 0, \quad dW^A / d\tau^C = 0, \quad j \in \{B, C\} \); \quad (R1)
3. \( dW^j / d\tau^j > 0, \quad j, \land j \in \{B, C\} \).

2.3. The bargaining structure

In the following sections we explore whether the efficiency frontier can be reached in a variety of specific bargaining environments. Common across these environments is the feature that negotiations proceed bilaterally and sequentially under MFN. As we noted in the Introduction, bilateral tariff bargaining under MFN is a staple of the GATT/WTO whose antecedents date back to the bilateral tariff bargaining procedures under the RTAA. Fig. 1 illustrates the basic structure of the sequence of bargaining for the governments of countries A, B and C. According to our economic model, exporters from countries B and C sell into A’s market, while exporters from country A sell into the markets of B and C, but B and C do not trade with each other. As a consequence, Fig. 1 depicts a sequence of bilateral MFN market access negotiations, first between A and B over the tariffs each controls and the transfer between them, and second between A and C over the tariffs each controls and the transfer between them.

3. Backward stealing and bargaining inefficiencies

According to Lemma 2, any point on the efficiency frontier must satisfy (R1), and under (R1) efficiency condition (4) implies:

\[
-W^j_{\tau^j} / W^A_{\tau^C} = 0 > -W^A_{\tau^A} / W^B_{\tau^A} > -W^j_{\tau} / W_{\tau^j} \quad \text{for} \quad j, \land j \in \{B, C\} \quad (5)
\]

where we use \( dr^j/d\tau^j = 1 \) for \( j \in \{B, C\} \). With \( \tau^j \) on the vertical axis and \( t^j \) on the horizontal axis, Fig. 2 depicts the “lens” implied by Eq. (5). As Fig. 2 illustrates, beginning from any efficient policy combination, the governments of country A and either of its trading partners can enjoy mutual gains – at the expense of the government of the third country – if A’s transfer to this trading partner is increased slightly above the efficient level (denoted \( \tau^k \)) and the trading partner’s tariff is reduced slightly below the efficient level (denoted \( \tau^k \)). We summarize this observation with:

**Lemma 3.** At any point on the efficiency frontier, and for \( j, \land j \in \{B, C\} \), it is possible to reduce \( \tau^j \) and increase \( t^j \) so as to increase \( W^A \) and \( W^j \) at the expense of \( W^B \).

The lens described in Lemma 3 is significant, because it signals the broad potential for a “backward-stealing" problem when governments negotiate bilaterally and sequentially, even when those negotiations are constrained to abide by MFN. In effect, with no change to its own tariff whatsoever, the government of country A can use its transfer policy to "pay“ one of its trading partners to liberalize and

---

9 One reason for the reciprocal nature of GATT/WTO tariff commitments is to increase compliance with the negotiated commitments. In particular, enforcement in the GATT/WTO is achieved primarily through the threat of withdrawal of negotiated tariff commitments (see Bagwell and Staiger, 2002, and Bown, 2004), a threat that would be unavailable to a country that was already on its reaction curve. While we abstract from such enforcement issues in our formal analysis here, they provide an additional reason for our below-the-reaction-curve curve focus.
generate a beneficial improvement in A’s terms of trade, all at the expense of the third-party country.

We now define the Basic Sequential MFN Game or, for short, the Basic Game. In stage 1, country A makes a take-it-or-leave-it proposal to B concerning tariff bindings (i.e., permissible upper bounds on applied tariffs) $\tau^A$ and $\tau^B$, as well as a transfer from A to B, $\tau^F$. Then, in stage 2, country A makes a take-it-or-leave-it proposal to C concerning bindings $\tau^A$ (with the stage-2 binding $\tau^B$ set no higher than its stage-1 level $\tau^B$) and $\tau^C$, as well as a transfer from A to C, $\tau^F$. The Basic Game has the following features:

Stage 1: A proposes $(\tau^A, \tau^B, \tau^F)$, which B accepts or rejects.
Stage 2: If B accepts, A proposes $(\tau^A \leq \tau^A, \tau^C, \tau^F)$, which C accepts or rejects. If B rejects, A proposes $(\tau^A, \tau^B, \tau^C)$, which C accepts or rejects.

The Basic Game is illustrated in Fig. 3. Here and throughout the paper, we assume that, subsequent to the conclusion of negotiations (e.g., after stage 2 of the Basic Game), each government sets its applied tariff unilaterally and simultaneously with the other governments subject to the constraint that it does not exceed its negotiated tariff binding(s).

We impose a global “stability” condition on tariff reaction curves to rule out the existence of “unstable” Nash equilibria. Denoting $j$'s best-response tariff function by $\tau^B_j(\cdot)$ for $j = \{A, B, C\}$ (and recalling that subscripts denote partial derivatives), this stability condition is recorded in:

$$1 > \left[ \frac{\partial \tau^B_j}{\partial \tau^A} + \frac{\partial \tau^B_j}{\partial \tau^B} \right] \frac{\tau^B_j}{\tau^A} + \frac{\partial \tau^B_j}{\partial \tau^C} \frac{\tau^C}{\tau^A} \left[ 1 - \tau^B_j \frac{\tau^B_j}{\tau^A} \right]$$ (A3)

In words (A3) ensures, for example, that binding A’s tariff below its reaction curve could not induce changes in the best-response tariffs of B and C which, together, would induce an even greater reduction in A’s best-response tariff. As with (A1) and (A2), we treat (A3) as a maintained assumption in what follows.

Analysis of the Basic Game is made complicated by the fact that governments negotiate tariff bindings, which only place upper limits on their applied tariffs. We therefore describe at this point only the subgame which is featured in our next result, namely, the subgame in which both B and C accept A’s proposals. When both B and C agree, we denote the tariffs applied under the (full) agreement by $\tau^A = \tau^A(\bar{\tau}, \Gamma)$, $\tau^B = \tau^B(\bar{\tau}, \Gamma)$, and $\tau^C = \tau^C(\bar{\tau}, \Gamma)$, where $\bar{\tau}$ denotes the vector of final tariff bindings $(\tau^A, \tau^B, \tau^C)$ and $\Gamma$ denotes the vector of transfers $(\tau^A, \tau^B, \tau^C)$. The tariffs $\tau^A$, $\tau^B$, and $\tau^C$ are defined by the first-order conditions

$$W^A + \lambda^A = 0; \quad W^B + \lambda^B = 0; \quad \text{and} \quad W^C + \lambda^C = 0$$ (6)

evaluated with $\bar{\tau}^B = \tau^B$ and $\bar{\tau}^C = \tau^C$, and where $\lambda^A$, $\lambda^B$, and $\lambda^C$ are the Lagrange multipliers on the constraints $\tau^A \leq \tau^A$, $\tau^B \leq \tau^B$, and $\tau^C \leq \tau^C$, respectively. By Eq. (6), $\lambda^A = \min\{\tau^A, \tau^B(\bar{\tau}, \tau^C), \tau^C\}$ is the applied tariff for A, B’s applied tariff is $\bar{\tau}^B = \min\{\tau^B, \tau^C(\bar{\tau}, \tau^A), \tau^A\}$, and C’s applied tariff is $\bar{\tau}^C = \min\{\tau^C, \tau^A(\bar{\tau}, \tau^B), \tau^B\}$. In this subgame, the payoffs for A, B, and C, are respectively $W^A(\bar{\tau}, \bar{\tau}^B, \bar{\tau}^C), w^B(\bar{\tau}, \bar{\tau}^A, \bar{\tau}^C), w^C(\bar{\tau}, \bar{\tau}^A, \bar{\tau}^B, \bar{\tau}^C)$.

Fig. 2. The backward-stealing lens for $j \in \{B, C\}$.

Fig. 3. The basic game.

---

11 In subsequent sections, we introduce notation for and discuss in detail the payoffs that are received when some or all of A’s proposals are rejected. For the present section, we need only note that the payoff for C when B accepts and C rejects is independent of the specific proposal that C chose to reject.

12 We assume that an interior Nash equilibrium exists in which each country trades in its “natural” direction, i.e., in the direction that would prevail absent tariffs. As Dixit (1987) observes, a satisfactory Nash equilibrium may exist as well. In the event that B and C reject A’s offer, we assume that the interior Nash equilibrium is played, with $j$ then receiving welfare $W^j$ for $j = \{A, B, C\}$.
We focus on subgame perfect equilibria (SGPE) of the Basic Game. We will say that the outcome is efficient (inefficient) when the payoffs correspond to a point on (off) the efficiency frontier. Under (A1), we are interested in points on the efficiency frontier where each country’s (applied) tariff is constrained to lie below its reaction curve. Clearly, achieving such a point as the outcome of the Basic Game requires that A reach agreement with both B and C, and at such a point each country’s applied tariff is then set equal to the level of its binding, or τ̃ = τf j for j ∈ {A, B, C}. From our description of the subgame just above, we must then have wA(τ̃, t̃ A) = wB(τ̃, t̃ B), wB(τ̃, t̃ B) = wC(τ̃, t̃ C), and wC(τ̃, t̃ C) = wA(τ̃, t̃ A). Hence, we ask whether there exists a SGPE of the Basic Game in which the associated choices of (τ, t) imply a triple (wA(τ, t A), wB(τ, t B), wC(τ, t C)) that is efficient. We now state:

**Proposition 1.** There does not exist a SGPE of the Basic Game in which the outcome is efficient.

This result may be seen as follows. Starting from stage-2 choices that would achieve the efficiency frontier, A and C can do better for themselves if C liberalizes further (i.e., reduces 1/τC). C’s import liberalization benefits A by increasing the price of A’s export good on world markets (i.e., by improving A’s terms of trade), and A can compensate C for Cs implied welfare loss with an increased transfer to C (i.e., increased 1/τC) while enjoying the gains from higher export prices against B. Hence, efficient outcomes are precluded by the backward-stealing problem identified in Lemma 3.

Finally, we observe that, while we have derived Proposition 1 in a take-it-or-leave-it bargaining context, it is clear from Fig. 2 (with j = C) that the proposition holds in more general bargaining environments as well, provided only that the stage-2 bargain between A and C is efficient (i.e., exhausts all feasible gains from cooperation in that stage) and therefore leads to a tangency between the indifference curves of A and C in Fig. 2.

### 4. Forward manipulation and bargaining inefficiencies

Backward stealing prevents efficient outcomes in the Basic Game analyzed in the previous section. Suppose, then, that a “security constraint” were introduced into the Basic Game, wherein the governments of countries A and C were prevented from reducing the welfare of B with their negotiations. (We postpone the question of how such a constraint might be maintained until Section 6.) Could governments achieve efficient outcomes in this augmented bargaining game? In this section, we show that the answer to this question is generally “No.” More specifically, we identify an incentive for “forward manipulation” that can keep governments from the efficiency frontier.

To accomplish this, we impose on the Basic Game the following security constraint: any agreement reached in stage-2 between A and C must leave B at least as well off as it would be if instead the negotiations between A and C ended in disagreement and only the stage-1 agreement were implemented. Under this security constraint, there certainly can be no backward stealing. Hence, we say that a stage-1 agreement between A and B is secure against backward stealing if and only if, following an agreement between A and B in stage 1, any agreement between A and C satisfies this security constraint.

In order to formalize this constraint, we must define B’s payoff in the subgame of the Basic Game where B accepts and C rejects. In this subgame, there is no transfer between A and C, and C does not agree to bind its tariff, while A and B agree to bind their tariffs and agree as well to a transfer between them. Hence, in this subgame, C selects its best-response tariff, τRC, to the tariffs applied by A and B under their agreement. We denote the tariffs applied by A and B under their agreement by τAC ∈ τAC(τA, τB, τB) and τBC ∈ τBC(τA, τB, τB), respectively. The tariffs τRC, τAC, and τBC satisfy the three first-order conditions

\[ W^B_C = 0; \quad W^A_C + \lambda^{A:C} = 0; \quad \text{and} \quad W^B_A + \lambda^{B:C} = 0 \quad (7) \]

evaluated with τ̃ = τB and t̃ = 0, and where λ^{A:C} and λ^{B:C} are the Lagrange multipliers on the constraints τAC ≤ τA and τBC ≤ τB, respectively. By Eq. (7), λ^{A:C} = min{τB, τAC(τA, τB, τB)} is A’s applied tariff, and B’s applied tariff is τBC = min{τB, τBC(τA, τB, τB)}. In this subgame, then, B receives WBC(τ̃, τ̃ B) = WBC(τAC, τBC, τCR, τB). We may now state the security constraint formally:

\[ W^B_C(τ̃, t̃) ≥ W^B_C(τ^A_C, τ^B_C, τ^B_R) \quad (8) \]

We next define the Secure-Contract Game. In stage 1 of this game, country A makes a take-it-or-leave-it proposal to B concerning bindings τA and τB, as well as a transfer from A to B, τB. Then, in stage 2, country A makes a take-it-or-leave-it proposal to C concerning bindings τA with the stage-2 binding τB set no higher than its stage-1 level τA, and τC, as well as a transfer from A to C, τC, subject to ensuring that any agreement reached in stage 1 is secure against backward stealing. The Secure-Contract Game has the following features:

**Stage 1:** A proposes (τA, τB, τB), which B accepts or rejects.

**Stage 2:** If B accepts, A proposes (τ̃B ≤ τA, τC, τC), where wB(τ̃, t̃B) ≥ wBC(τ̃B, τA, τC), which C accepts or rejects. If B rejects, A proposes (τ̃B, τ̃C, τ̃C), which C accepts or rejects.

We introduce a second “stability” condition for the best-response tariff functions:

Each country’s best-response tariff function everywhere satisfies “terms-of-trade stability,” i.e., for i ≠ j ≠ k ∈ {A, B, C},

\[ \left( \frac{∂ τ_i}{∂ τ_j} \right) \times \left( \frac{∂ τ_k}{∂ τ_i} \right) + \frac{∂ τ_k}{∂ τ_j} + \frac{τ_j + τ_k + τ_{jR} + τ_{kR} + τ_{jR} + τ_{kR}}{τ_i + τ_j + τ_k + τ_{jR} + τ_{kR}} > 0 \quad (A4) \]

In words (A4) ensures, for example, that the drop in the market-clearing terms of trade implied by a reduction in B’s tariff below its reaction curve would not be completely reversed by the best-response tariff adjustments of A and C that B’s tariff reduction would induce.

To characterize the SGPE of the Secure-Contract Game, it is useful to begin by considering the disagreement welfare levels in this game for the governments of countries B and C, in the event that A reaches agreement with the other trading partner.

For B, this disagreement welfare is determined by the equilibrium of the stage-2 subgame between A and C that follows stage-1 disagreement between A and B. In this subgame there is no transfer between A and B, and B does not agree to bind its tariff, while A and C agree to bind their tariffs and agree as well to a transfer between them. Hence, in this subgame, B selects its best-response tariff, τBR, to the tariffs applied by A and C under their agreement. We denote the tariffs applied by A and C under their agreement by τAC ≡ τAC(τA, τC, τC) and τBC ≡ τBC(τA, τC, τC), respectively. The three tariffs τBR, τAC, and τBC are defined by the three first-order conditions

\[ W^B_A = 0; \quad W^A_C + \lambda^{A:B} = 0; \quad \text{and} \quad W^B_C + \lambda^{B:C} = 0 \quad (9) \]

evaluated with τ̃ = τB and t̃ = 0, and where λ^{A:B} and λ^{B:C} are the Lagrange multipliers on the constraints τAC ≤ τA and τBC ≤ τB.

---

11 When it is clear from context, we let τ̃ denote j’s best-response tariff to the applied tariffs of A and |j| when t̃ = 0 for j = (B, C).
respectively. By Eq. (9), \( T^{AB}_{\text{eff}} = \min\{ T^{A}, T^{B}_{\text{eff}}(T^{AB}; T^{B}_{\text{eff}}(T^{C}); T^{C})\} \) is A’s applied tariff, and \( T^{CB}_{\text{eff}} = \min\{ T^{C}, T^{B}_{\text{eff}}(T^{AB}; T^{C}_{\text{eff}}; T^{C})\} \) is C’s applied tariff. In this subgame, then, B’s payoff is \( w^{B}(T^{A}; T^{C}_{\text{eff}}; T^{C}) = W^{B}_{\text{ref}}(T^{A}; T^{C}_{\text{eff}}; T^{C}) \). Letting A’s equilibrium stage-2 proposal to C in this subgame be denoted by \( T^{AB} = \min\{ T^{A}, T^{C}_{\text{eff}}(T^{AB}; T^{B}_{\text{eff}}(T^{C}); T^{C})\} \) and observing that C will accept the equilibrium proposal, B’s disagreement welfare in its stage-1 negotiation with A is then given by \( w^{B}_{\text{ref}}(T^{AB}; T^{C}_{\text{eff}}; T^{C}) \). For future reference, we denote B’s disagreement welfare by \( W^{B}_{\text{ref}} \).

Importantly, as viewed from stage 1, \( W^{B}_{\text{ref}} \) is tied down by the requirement of subgame perfection, and A therefore has no means by which to (credibly) manipulate B’s disagreement welfare to its own advantage. Circumstances are different, however, for the government of country C. Its disagreement welfare is \( W^{C}_{\text{ref}}(T^{A}; T^{B}; T^{C}) = W^{C}_{\text{ref}}(T^{A}; T^{B}; T^{C}; T^{C}) \) for any such \( (T^{A}; T^{B}; T^{C}) \) that together with \( T^{C} \) fail to drive C to autarky, \( W^{C}_{\text{ref}}(T^{A}; T^{B}; T^{C}) \) is strictly decreasing in \( T^{A} \), strictly increasing in \( T^{B} \), and independent of \( T^{C} \).

In effect, with disagreement placing C on its reaction curve and for any \( T^{A} < T^{B} < T^{C} \), the implied world price \( P^{B}(T^{A}; T^{B}; T^{C}) \) is lower than \( T^{B} \). Thus, \( W^{C}_{\text{ref}}(T^{A}; T^{B}; T^{C}) \) is driven to the minimum (autarky) level, which we denote by \( W^{C}_{\text{ref}} \).

We wish to explore the conditions under which the SGPE of the Secure-Contract Game (SCG) is in an equilibrium. Consider any stage-1 proposal \( (T^{A}, T^{B}, T^{C}) \) that would induce the Secure-Contract Game (SCG) to place a point on its efficiency frontier \( (W^{BE}, W^{BE}, W^{CE}) \) for which \( W^{CE} > \min\{W^{BE}, W^{CE}\} \). This proposal must satisfy

\[
W^{B}(T^{A}; T^{B}; T^{C}) = W^{BE} \geq W^{BD} \tag{10a}
\]

and

\[
W^{C}(T^{A}; T^{B}; T^{C}) = W^{CE} \tag{10b}
\]

The first part of condition (10a) states that the security constraint must hold with equality: if it held with strict inequality there would be room for backward unraveling according to Proposition 1. The second part of condition (10a) states that B must receive at least its disagreement welfare level. And the condition in (10b) must hold, since otherwise either C would reject A’s stage-2 offer (if \( T^{C} < T^{A} \)) or A could reduce the transfer to C in its stage-2 offer and do better (if \( T^{C} > T^{A} \)). In addition, this proposal must satisfy

\[
t^{A} < T^{B} \leq \left( T^{B}_{\text{eff}}(T^{C}; T^{C}) \right) \quad \text{and} \quad \left( T^{C}_{\text{eff}}(T^{AB}; T^{B}; T^{C}) \right) < T^{B}
\]

Otherwise, the proposal could not induce a point on the efficiency frontier.\(^{15}\) We prove in the Appendix:

**Proposition 2.** Under (A4), there does not exist (generically) a SGPE of the Secure-Contract Game in which the outcome is efficient and \( W^{CE} > \min\{W^{BE}, W^{CE}\} \).

Intuitively, beginning from proposals that efficiently deliver \( W^{BE} \) and \( W^{CE} \), if \( W^{CE} > \min\{W^{BE}, W^{CE}\} \) then Eq. (11) and (A4) together ensure that there exists (generically) an adjustment in \( T^{A} \) or \( T^{B} \) which reduces C’s payoff and, along with adjustments in \( T^{B} \), \( T^{A}, \) and \( T^{C} \), improves A’s payoff.

Evidently, efficiency in the Secure-Contract Game requires that \( W^{CE} = \min\{W^{BE}, W^{CE}\} \) and this requirement places rather extreme demands on the environment within which the Secure-Contract Game can deliver governments to the efficiency frontier. In particular, with \( T^{B} \) tied down by the combination of policies to which A must navigate to achieve this outcome, it must then be feasible for A to position \( T^{A} \) so that C faces autarky if it rejects A’s stage-2 proposal.\(^{16}\) If this cannot be accomplished with a choice of \( T^{A} \), because for example accomplishing this would require A to set an applied tariff \( T^{A} \) that was above its best-response level, then A cannot both give C its minimal payoff \( W^{CE} \) and achieve efficiency, and in this case Proposition 2 indicates that A will find it desirable to sacrifice efficiency in order to lower C’s payoff toward \( W^{CE} \). More generally, Proposition 2 suggests that governments will achieve efficient bargaining outcomes in the Secure-Contract Game under some special circumstances, but under many plausible circumstances the outcome of the Secure-Contract Game is inefficient.

The source of inefficiency identified in Proposition 2 arises from A’s desire to use its stage-1 negotiations with B to position itself more favorably for stage-2 negotiations with C by sticking C with a less-favorable disagreement point. Prop. 4 illustrates with \( T^{C} \) on the vertical axis and \( F^{B} \) on the horizontal axis, we consider a stage-1 proposal under which it would be “feasible” to achieve efficiency, i.e., a stage-1 proposal \( (T^{A}, T^{B}, T^{C}) \) that would induce in the Secure-Contract Game the efficient point \( (W^{BE}, W^{BE}, W^{CE}) \) with \( (W^{BE} = W^{BD}) \) and \( W^{CE} > \min\{W^{BE}, W^{CE}\} \). We have observed that attention may be restricted to stage-1 proposals that satisfy Eq. (11), and for purposes of illustration we assume that \( T^{A} < T^{B} < T^{C} \) and \( T^{B} > T^{C} \). According to Proposition 2, A would not choose \( T^{B} \) and \( T^{C} \) if it chooses \( T^{A} \). We wish to illustrate in this case that it is always possible for A to raise its welfare by proposing \( T^{A} > T^{B} \) and \( T^{C} > T^{B} \). To understand why, let us fix \( T^{A} = T^{B} \) and consider varying the proposed \( T^{C} \) and \( T^{B} \) around \( T^{A} \) and \( T^{B} \), respectively, as so hold \( W^{B}(T^{A}; T^{B}; T^{C}) \) fixed at \( W^{BE} \).

\(^{15}\) To see this, suppose that Eq. (11) is violated so that \( T^{A} > T^{B} > T^{C} \) and \( T^{B} > T^{C} \). In this case, if C were to disagree in stage 2, A’s stage-1 proposal would implement the Nash tariff equilibrium under the transfer \( T^{B} \). Moreover, to satisfy Eq. (10a), the implied world price under the stage-1 proposal must be the same as the world price implied by the full agreement: if the world price were to move against B in the full agreement, B would be hurt as long as it remained on its reaction curve, and hurt further once it was held below its reaction curve by \( T^{B} \), and so Eq. (10a) could not hold for world price movements in this direction; if the world price were to move in B’s favor, B would benefit as long as it remained on its reaction curve, and could be hurt by further world price movements in this direction once it was forced below its reaction curve by \( T^{B} \), but satisfying Eq. (10a) in this fashion would be inconsistent with efficiency and (A1) and (A2). So the (Nash) world price implied under the stage-1 proposal is also the world price implied under the full agreement. Finally, according to (A1), we must have B’s applied tariff rise from its (Nash) level under the stage-1 agreement to the efficient level \( T^{B} \). This implies, then, that the transition from the stage-1 agreement to the full agreement describes a movement from the Nash tariff equilibria (under the transfer \( T^{B} \)) to a point on the efficiency frontier, in which (i) the world price remains unaltered, (ii) B’s tariff is raised, and therefore (iii) B’s trade volume is diminished. But it is easy to demonstrate that achieving a point on the efficiency frontier at the Nash world price is inconsistent with a drop in B’s trade volume.

\(^{16}\) More specifically, achieving \( W^{CE} = \min\{W^{BE}, W^{CE}\} \) requires that C face its autarky terms-of-trade \( T^{A} \) in the event that it rejects A’s stage-2 proposal, and therefore, by (A3) and (A4) and maintaining our focus on interior Nash equilibria, a stage-1 proposal that achieves efficiency requires that \( T^{A} < T^{A}(T^{B}, T^{C}) \). As a consequence, \( T^{A} \) and \( T^{B} \) must be consistent with efficiency and satisfy \( T^{A} = T^{A}(T^{B}, T^{C}) = T^{B} \) by Eq. (10a).
security constraint, B will accept any such proposal. Further, let the stage-2 proposal of \( \{ \tau^A, \tau^C, \tau^B \} \) that is associated with any \( \{ \tau^A, \tau^B, \tau^C \} \) maximize \( W^A(\tau, \tau^A) \) subject to \( W^B(\tau, \tau^B) = w^B(\tau^A, \tau^B, \tau^C) \) and \( W^C(\tau, \tau^C) = w^C(\tau^A, \tau^B, \tau^C) \) subject to \( \tau^C \leq \tilde{\tau}^C \). Clearly, if \( \tau^C \) will accept the stage-2 proposal, and so by construction the welfare level for A associated with \( W^A(\tau, \tau^A) \) attainable in the Secure-Contract Game. In effect, then, with \( \tau^B \) fixed at \( \tilde{\tau}^B \), any change in \( \tau^A \) implies by this construction an associated change in \( \{ \tau^A, \tau^C, \tau^B \} \). Figure 4 depicts the welfare consequences of these changes.

Consider then, the indifference curves associated with \( W^A(\tau, \tau^A) \), \( W^B(\tau, \tau^B) \) and \( W^C(\tau, \tau^C) \) under this construction which pass through the efficient point \( (\tau^A, \tau^B, \tau^C) \) in Figure 4. C’s indifference curve is horizontal through this point, since \( W^C(\tau, \tau^C) = w^C(\tau^A, \tau^B, \tau^C) \) is strictly decreasing in \( \tau^A \) and independent of \( \tau^B \) by Lemma 4. B’s indifference curve through this point could have positive or negative slope (it is depicted in the figure with positive slope) but, importantly, it cannot be horizontal, since \( W^B(\tau, \tau^B) = w^B(\tau^A, \tau^B, \tau^C) \) is strictly increasing in \( \tau^B \). Therefore, the indifference curves associated with \( W^B(\tau, \tau^B) \) and \( W^C(\tau, \tau^C) \) are not tangent to each other as they pass through the efficient point \( (\tau^A, \tau^B, \tau^C) \) as depicted in the figure with positive slope, but, importantly, it cannot be horizontal, since \( W^B(\tau, \tau^B) = w^B(\tau^A, \tau^B, \tau^C) \) is strictly increasing in \( \tau^B \). Therefore, the indifference curves associated with \( W^B(\tau, \tau^B) \) and \( W^C(\tau, \tau^C) \) are not tangent to each other as they pass through the efficient point \( (\tau^A, \tau^B, \tau^C) \) in Figure 4, and as a consequence, by efficiency condition (4) A’s indifference curve cannot be tangent to either B’s or C’s indifference curve at this point. Moreover, A’s indifference curve must be flatter than B’s at this point, since otherwise a “downward” lens would exist between the indifference curves of A and B into which A and B could move and all three governments would gain, contradicting the efficiency of this point. Therefore, as Figure 4 depicts, there is an “upward” lens created by the indifference curves of A and B at this point, implying that A can gain by raising \( \tau^A \) above \( \tilde{\tau}^A \) and adjusting \( \tilde{\tau}^B \) to maintain B’s welfare. A’s gain from this maneuver comes at the expense of C’s welfare (and multilateral efficiency).17

We observe that this logic is related to a concern about the “foot-dragging” maneuver for handling free-riders as this maneuver was described in the Introduction. According to this concern, country A might be induced under MFN to offer “too little” in the way of trade liberalization to its early negotiating partners, in order to maintain its bargaining position for later negotiations. Proposition 2 provides a formal justification for this concern, and Figure 4 illustrates the foot-dragging incentive to maintain \( \tau^A \) above its efficient level.

In summary, Proposition 2 indicates that the Secure-Contract Game can deliver efficient bargaining outcomes only in a very limited set of circumstances. In the circumstances where inefficiency arises, this inefficiency is associated with forward manipulation. In the next section, we consider how this new source of inefficiency might be handled.

5. Preventing forward manipulation through GATT/WTO rules

One way to correct the inefficiency associated with forward manipulation is to eliminate the possibility of forward manipulation itself. In principle, this might be achieved by introducing renegotiation opportunities, provided that these renegotiation opportunities are sufficiently “sweeping” so that they separate C’s disagreement payoff from the stage-1 determination of \( \{ \tau^A, \tau^B, \tau^C \} \). In fact, during the United States experience with bilateral MFN tariff bargaining under the RTAA, the threat to withdraw or modify a concession granted to an early negotiating partner was seen as providing an important “off equilibrium path” means of maintaining bargaining leverage for later negotiations with other countries (see Tasca, 1938, p. 146), and it may have reduced to some extent the reliance on foot-dragging for this purpose.

The GATT/WTO explicitly provides for renegotiation. This is true both within a multilateral round of negotiation, when agreements reached between negotiating pairs early in the round are viewed as tentative and may be revisited if subsequent negotiations with other partners do not go as expected (e.g., Jackson, 1969, p. 220), and it is also true outside of multilateral rounds, where explicit renegotiations of previous agreements are permitted (e.g., Jackson, 1969, pp. 229–238). Just how sweeping these renegotiating opportunities are is a question of degree, and presumably depends on circumstances.

In this section, we consider whether introducing sweeping renegotiation possibilities into the Secure-Contract Game can solve the forward-manipulation problem and lead (in the presence of the security constraint) to efficient outcomes. We first describe the novel features of the Contract Renegotiation Game:

Stage 1: A proposes \( \{ \tau^A, \tau^B, \tau^C \} \), which B accepts or rejects.
Stage 2: If B accepts, A proposes \( \{ \tau^A \leq \tau^B, \tau^C, \tau^b \} \), where \( w^B(\tau, \tau^C) \geq w^B(\tau^A, \tau^B, \tau^C) \), which C accepts or rejects. If B rejects, A proposes \( \{ \tau^A, \tau^C, \tau^b \} \), which C accepts or rejects.
Stage 3: If B accepts in stage 1 and C rejects in stage 2, then A proposes \( \{ \tau^A, \tau^B, \tau^C \} \), which B accepts or rejects.

The Contract Renegotiation Game is illustrated in Figure 5. Notice that we have modeled the renegotiation stage 3 as arising only if B accepts A’s original proposal in stage-1 (we have also modeled renegotiation as arising only if C rejects in stage 2, but this is not crucial).18 With this modeling choice, we attempt to capture the relative ease with which earlier “tentative” agreements can be adjusted within GATT/WTO rounds and/or earlier commitments can be renegotiated under GATT Article XXVIII between rounds.

In the Secure-Contract Game, the source of inefficiency can be traced to the problem of forward manipulation, whereby A’s stage-1 proposal to B influences C’s disagreement payoff in its stage-2 negotiations with A. In the Contract Renegotiation Game, this linkage has been curtailed, but it has not been eliminated. To see this, note that if B accepts A’s stage-1 proposal, and if C then rejects in stage 2, C’s disagreement payoff will be determined by the renegotiation between A and B in stage 3. In this case, the details of A’s stage-1 proposal

\[17 \text{As noted, Figure 4 illustrates the case where } \tilde{\tau}^B = \tilde{\tau}^B \text{ and } \tilde{\tau}^C = \tilde{\tau}^C \text{. If there exists a } \{ \tilde{\tau}^A, \tilde{\tau}^B \} \text{ under which it is feasible to efficiently deliver } \tilde{\tau}^B = \tilde{\tau}^B \text{ and } \tilde{\tau}^C = \tilde{\tau}^C \text{ in the Secure-Contract Game, then at this point there is no lens between A and B. For example, in the case where } \tilde{\tau}^A < \tilde{\tau}^B < \tilde{\tau}^C \text{ and } \tilde{\tau}^C = \tilde{\tau}^C \text{, a small change in } \tilde{\tau}^B \text{ and an accompanying change in } \tilde{\tau}^C \text{ that left B indifferent would leave C indifferent as well (C would be indifferent to any small change in } \tilde{\tau}^A \text{ and } \tilde{\tau}^B \text{ by the first-order condition that defines } \tilde{\tau}^C \text{), and so by efficiency condition (4) A’s indifference curve must be tangent to B’s indifference curve as well in this case.}

\[18 \text{As Figure 5 illustrates, in the event that (i) B rejects in stage 1 and C rejects in stage 2, or (ii) B accepts in stage 1, C rejects in stage 2 and then B rejects in stage 3, we assume that the players adopt their Nash equilibrium strategies and earn the corresponding Nash payoffs.} \]
C’s payoff in this subgame is t concerns the possibility that A might choose to confront a negotiating down by subgame perfection, and A therefore cannot manipulate these subgame as disagreement between A and B. This subgame is identical to that in the stage-2 subgame between A and C that follows stage-1 country B, this disagreement welfare is determined by the equilibrium agreement with the other trading partner. For the government of governments of countries B and C, in the event that A reaches the ef

Fig. 5. The contract renegotiation game.

proposal to B are immaterial for C’s disagreement payoff in its stage-2 negotiations with A, provided only that B accepts A’s stage-1 proposal and therefore “locks in” a stage-3 renegotiation opportunity with A should C disagree in stage 2. However, if B does not accept A’s stage-1 proposal, then B has no renegotiation rights with A in stage 3, and so in this case C’s disagreement payoff in its stage-2 negotiations with A will be its Nash payoff. As a consequence, in the Contract Renegotiation Game the possibility of forward manipulation has been reduced to the question of “bypass”: Might A choose to make an unacceptable proposal to B in stage 1 in order to bypass B and negotiate with C against a Nash disagreement payoff? The question of bypass can also be seen to arise with C: Having made a stage-1 proposal that was accepted by B, might A choose to make an unacceptable proposal to C in stage 2 in order to bypass C and renegotiate with B against a Nash disagreement payoff in stage 3?

To ensure that the bypass problem does not prevent governments from reaching the efficiency frontier in the Contract Renegotiation Game, an additional condition is needed. To state this condition, consider the particular disagreement welfare levels in this game for the governments of countries B and C, in the event that A reaches agreement with the other trading partner. For the government of country B, this disagreement welfare is determined by the equilibrium of the stage-2 subgame between A and C that follows stage-1 disagreement between A and B. This subgame is identical to that in the Secure-Contract Game, and we previously recorded B’s payoff in this subgame as \( w_{BC}(t^A, t^B, t^C) \equiv W_B(t^A, t^B, t^C) \), which we denoted by \( \tilde{W}_B \). For the government of country C, this disagreement welfare is determined by the equilibrium of the stage-3 (renegotiation) subgame between A and B that follows stage-1 disagreement between A and B and stage-2 disagreement between A and C. Denoting A’s equilibrium proposal in this subgame by \( \tilde{\tau}_A \), C’s payoff in this subgame is \( w_{CD}(t^A, \tilde{\tau}_A, t^B, t^C) \equiv W_C(t^A, \tilde{\tau}_A, t^B, t^C, t^C) \equiv 0 \), which for future reference we denote by \( \tilde{W}_C \).

Importantly, as viewed from stage 1 (stage 2), \( \tilde{W}_B (\tilde{W}_C) \) is tied down by subgame perfection, and A therefore cannot manipulate these disagreement payoff levels to its own advantage. Bypass, however, concerns the possibility that A might choose to confront a negotiating partner with the Nash disagreement payoff \( \tilde{W}^{ij} \) rather than \( \tilde{W}^D \) for \( j = \{B, C\} \). We now state a condition to rule out the bypass problem:

\[
\tilde{W}^{BD} \leq W^{BN}; \quad \tilde{W}^{CD} \leq W^{CN}
\]  

(A5)

Under (A5), the minimal disagreement payoffs for B and C in the Contract Renegotiation Game are given by \( \tilde{W}^{BD} \) and \( \tilde{W}^{CD} \), respectively, and each of these disagreement payoffs is achieved only when A reaches an initial agreement with the other trading partner.\(^{19}\) Condition (A5) essentially requires that B and C not be too asymmetric with A when they participate in a multilateral Nash tariff war.\(^{20}\) We prove in the Appendix:

**Proposition 3.** Under (A5), forward manipulation does not impede the attainment of the efficiency frontier in the Contract Renegotiation Game.

When viewed in light of Proposition 2, Proposition 3 suggests that renegotiation provisions such as those provided in the GATT/WTO can

---

\(^{19}\) It might be wondered why bypass does not arise in the Secure-Contract Game. The reason is that, after B accepts, A can pin C at a welfare level that is lower than \( W_A \), since no further negotiations follow, and so A has no incentive to bypass B.\(^{20}\) Formally, for \( j = \{B, C\} \), and with \( r \equiv 0 \) or \( r' \), consider the three tariffs defined by \( \tilde{W}_j = 0 \) for \( j = \{A, B, C\} \). With this definition in hand, it may now be seen that, if \( A \) or \( j \) are symmetric participants in a multilateral Nash tariff war, neither succeeds in moving the terms of trade in its favor relative to the terms of trade that would obtain if each chose its tariff “without terms-of-trade considerations in mind,” i.e., so as to solve \( \tilde{W}_i = 0 \) for \( i = \{A, B, C\} \). Intuitively, with country \( j \) positioned on its reaction curve, when countries \( A \) and \( j \) are symmetric participants in a multilateral Nash tariff war, neither succeeds in moving the terms of trade in its favor relative to the terms of trade that would obtain if each chose its tariff “without terms-of-trade considerations in mind,” i.e., so as to solve \( \tilde{W}_i = 0 \) for \( i = \{A, B, C\} \). With this definition in hand, it may now be seen that, if \( A \) or \( j \) are symmetric participants in a multilateral Nash tariff war, then a bilateral agreement between them could achieve \( \tilde{W}_j = 0 \) for \( A \) and \( \tilde{W}_j = 0 \) for \( j \) while preserving the Nash terms of trade, thereby preserving as well the welfare of \( j \); but from here A could do better yet with an alternative proposal that worsened \( j \)’s terms of trade below the Nash terms of trade and compensated \( j \) with a higher transfer \( r' \), and the lower-than-Nash terms of trade under this alternative proposal would leave \( j \) with a lower-than-Nash welfare (for the proof that a government positioned on its tariff reaction curve experiences welfare changes that are the same sign as changes in its terms of trade, see the proof of Lemma 4 in the Appendix). Arguing in this fashion, it can be seen that (A5) holds if B and C are not too asymmetric with A when they participate in a multilateral Nash tariff war.
alleviate the efficiency costs associated with forward manipulation, in the sense that efficient bargaining outcomes may be anticipated in a wider set of circumstances, at least so long as these provisions allow for sufficiently “sweeping” renegotiation opportunities as we have modeled them here. Still, as [A5] indicates, as a general solution to forward manipulation, renegotiation has its limits, as it may introduce a bypass problem into negotiations in some circumstances (i.e., in the circumstances where [A5] is violated).

6. Preventing backward stealing through GATT/WTO rules

We now reconsider and interpret the security constraint imposed in Section 4 and maintained throughout Section 5 (i.e., the requirement that \( W^B(\tau, F) \geq W^B(\tau^*, \tau^B, F^B) \)). As suggested by our “terms-of-trade manipulation” interpretation of the backward-stealing problem offered in Section 3, there is a link between this security constraint and a requirement that stage-2 negotiations between A and C leave unaltered the terms-of-trade implied as a result of stage-1 negotiations. This link is suggestive of a role for GATT/WTO rules in this regard, because as we have shown elsewhere (Bagwell and Staiger, 1999, 2002) the GATT/WTO norm of reciprocity can be interpreted as fixing the terms of trade, in the sense that “reciprocal” changes in market access, i.e., changes that preserve the balance of market access rights and obligations, leave the terms of trade unaltered.21 Guided by this suggestion, we ask: Is there something in GATT/WTO rules that might work along the lines of such a security constraint?

To answer this question within our bargaining games, we focus on B’s opportunity to respond to subsequent negotiations between A and C that upset the original balance of B’s market access rights and obligations in a way that is unfavorable to B (i.e., that worsen B’s terms of trade from the level implied by B’s agreement with A). In this circumstance, we consider the possibility that B has the opportunity to respond with a tariff increase above its bound level that restores this original balance (i.e., that restores the original terms of trade). This opportunity can be seen to exist in the GATT/WTO within multilateral rounds of negotiation, in the sense that governments may choose to modify or withdraw tentatively offered concessions when imbalances arise at the conclusion of a round of GATT/WTO negotiations. And it can be seen to exist as well between rounds of multilateral negotiations, where governments have the opportunity to seek redress under the “non-violation nullification-or-impairment” provisio

The WTO-Contract Game is illustrated in Fig. 6. As compared to the Contract Renegotiation Game, the WTO-Contract Game displays two differences. First, in stage 2, the security constraint is no longer imposed. And second, immediately after stage 2 (in the new stage 3), B’s non-violation nullification-or-impairment response is inserted, permitting B to choose to increase its tariff binding if this is required to prevent A’s stage-2 agreement with C from eroding B’s terms of trade.23

We seek conditions under which any SGPE of the WTO-Contract Game will achieve a point on the efficiency frontier. We focus on the feasibility of efficiently delivering \( \bar{W}^B \) to B and \( \bar{W}^CD \) to C under (A5), since if this is feasible in the WTO-Contract Game then A will surely make proposals that implement it. We say that it is feasible in the WTO-Contract Game to efficiently deliver \( \bar{W}^B \) and \( \bar{W}^CD \) if and only if there exists a triple \( \{\tau^M, \tau^B, \tau^D\} \) such that the outcome of the WTO-Contract Game is efficient, satisfies (A1) and (A2), and gives B the payoff \( \bar{W}^B \) and C the payoff \( \bar{W}^CD \) when the stage-1 proposal is \( \{\tau^M, \tau^B, \tau^D\} \).

Further progress can be made by considering a particular combina


22 Formally, we may define the market access that one country affords to a second by the first country’s volume of import demand for the exports of the second country at a given world price. Hence, the market access that countries B and C each afford to A at a given world price \( \tau^p \) is defined by their respective import demands at that world price: \( MA^B(\tau^p, \tau^B, \tau^C) \equiv MA^C(\tau^p, \tau^B, \tau^C) \). Similarly, the market access that country A affords to country j is A’s residual import demand for j’s exports – the other country’s export supply to A has been netted out – at a given world price: \( MA^A(\tau^p, \tau^B, \tau^C) \equiv MA^A(p^j, \tau^B, \tau^C) \). Finally, we define the balance of market access rights and obligations between A and j that is implied by a vector of negotiated tariffs and transfers at a given world price \( \tau^p \) as \( MA^A(\tau^p, \tau^B, \tau^C) \equiv MA^A(p^j, \tau^B, \tau^C) \). For simplicity, we have omitted tariffs j under (A5), when all tariffs are evaluated at their applied levels. With this definition, it is then straightforward to show that two vectors of tariffs imply the same balance of market access rights and obligations if and only if they imply the same terms of trade. For further discussion of the relationship between the balance of market access rights and obligations and the terms of trade, see Bagwell and Staiger (2002).

23 We have abstracted from a number of the legal/institutional elements associated with non-violation complaints in the GATT/WTO (see Bagwell et al., 2002, for a recent discussion of non-violation complaints in the GATT/WTO). Among them is the notion of what level of market access B could reasonably anticipate it had attained in its stage-1 negotiation with A. Arguably, as the WTO is a forum for bilateral negotiations, it would be unreasonable for B not to anticipate that countries A and C might engage in subsequent negotiations. But these subsequent negotiations may be structured in a variety of ways, some of which could potentially have large adverse impacts on B’s interests. If the GATT/WTO norm of reciprocity is seen to define what B can reasonably anticipate concerning the outcome of A’s subsequent negotiations with C, then it may be concluded that the anticipated stage-2 negotiations will leave B unaffected, and hence the level of market access that country B can reasonably anticipate as a result of its stage-1 negotiations with country A is simply that which is implied by its stage-1 negotiations. Notice that, according to this argument, reciprocity is not being imposed as an additional restriction on the outcome of stage-2 agreements. Instead, reciprocity is introduced as a negotiating norm: if a bilateral negotiation does not satisfy this norm, then the parties to the negotiation may be vulnerable to claims of nullification or impairment by a third party, if the third party had previously negotiated a market access agreement with one of them. In this regard, Hudec (1990, pp. 23–24) notes that the designers of GATT added nullification-or-impairment provisions precisely out of a concern for maintaining reciprocity established by negotiated market access agreements.
j = (A, B, C), and it can be shown that politically optimal tariffs achieve the efficiency frontier defined by Eq. (4).

Intuitively, politically optimal tariffs are efficient, because the incentive of governments to manipulate the terms of trade with their tariff choices is the source of international inefficiency in the Nash equilibrium and politically optimal tariffs do not reflect this incentive. We denote by τ^po the vector of politically optimal tariffs τ^po for j = (A, B, C) that, along with associated transfers t^po, efficiently deliver W^RD and W^CD, and let P^po = P(τ^po).

By the first-order conditions that define them, politically optimal tariffs satisfy (A1) and (A2). Suppose, then, that there exists a τ^po ≥ τ^po such that the stage-1 proposal {τ^A = τ^Ap0, τ^B = τ^Bp0, τ^C = τ^Cp0} satisfies the condition P^0(τ^A, τ^B, τ^C) = P^po. If A were to make this proposal in stage-1 and B accepted, then the existence of τ^1 which defines B's stage-3 non-violation right in response to a stage-2 agreement reached between A and C is defined implicitly by P^0(τ^A, τ^B, τ^C) = P^po. If C were to accept this proposal, then τ^1 is in the range of τ^C which defines C's stage-1 non-violation right. Continuing in this way would select τ^B = τ^Bp0, and in this way W^RD and W^CD would be delivered efficiently with politically optimal policies. Hence, we may conclude that, provided there is no alternative stage-2 proposal which would be preferred by A, the existence of the stage-1 proposal described above is sufficient to ensure that it is feasible in the WTO Contract Game to efficiently deliver W^RD and W^CD, and therefore that the outcome of the WTO Contract Game will be efficient.

Consider, then, the possibility that A might deviate to an alternative stage-2 proposal. Observe first that politically optimal tariffs exhibit the special feature that no government would desire a different trade volume if this possibility were offered to it at fixed terms of trade (this is what τ^W = 0 for j = (A, B, C) means). As a consequence, A could not do better under a deviant stage-2 proposal that satisfied P^0(τ^A, τ^B, τ^C) = P^po in light of B's stage-3 response, i.e., under a deviant stage-2 proposal that implies τ^BNV = [τ^B, τ^C, (τ^A, τ^B, τ^C)]]. There are then two other possibilities.

A first possibility is that A could deviate to a stage-2 proposal implying τ^BNV < τ^po, in which case τ^B < τ^BNV and τ^C < τ^BNV. But even if this proposal provided A and C with their ideal trade volumes (implying W^B = 0 for j = (A, C)) at the new terms of trade, the decline in A's terms of trade implied by P^0(τ^A, τ^B, τ^C) < P^po ensures that A cannot gain under such a deviation. The remaining possibility is that A could deviate to a stage-2 proposal implying τ^BNV > τ^po, in which case τ^B > τ^BNV, C > τ^BNV, and P^0(τ^A, τ^B, τ^C) < P^po. The potential for A to gain from this kind of deviation arises because, with τ^B = τ^Bp0 already determined in stage 1, A might achieve higher welfare with tariff levels for itself and C which placed B on its tariff reaction curve than with politically optimal tariffs. Letting (τ^Bp0 + τ^Bp0, + τ^Cp0) denote the tariff choices that maximize W^A(τ^A, τ^B, τ^C, τ^A = τ^Bp0 + τ^Cp0) while delivering W^CD to C, this potential is ruled out by:

W^A(τ^po, τ^p0, τ^Cp0, τ^A = τ^Bp0 + τ^Cp0) ≥ W^A(τ^A, τ^B, τ^C, τ^A = τ^Bp0 + τ^Cp0 + τ^po)

(6A)

If (6A) is violated, then by negotiating with C, A could "win the tariff war" with B (i.e., with the transfer to B, τ^po, paid in either case, A could do better under non-cooperative tariff interaction with B than
under the politically optimal tariffs. Assumption (A6) effectively rules out this extreme degree of asymmetry, by requiring that B is not so small that A could win the tariff war. As there is no alternative stage-2 proposal which would be preferred by A, we may now state:

**Lemma 5.** Under (A5) and (A6), in any SGPE of the WTO-Contract Game, the outcome is efficient if there exists a $\tau_p^A \geq \tau_p^B$ and a $\tau_p^B \leq \tau_p^A$ such that the stage-1 proposal $(\tau^B = \tau_p^B, \tau^A = \tau_p^A, \tau^t = \tau_p^A)$ satisfies

$$ P^B(\tau^A, \tau^B, \tau^t) = P^A(\tau^A). $$

Lemma 5 provides a sufficient condition for efficient outcomes in the WTO-Contract Game. To interpret the circumstances under which this condition is met, we note that under (A3) and (A4), the highest value of $P^B(\tau^A, \tau^B, \tau^t)$ consistent with $\tau^A \geq \tau_p^A$ and $\tau^B \leq \tau^A$ is achieved at $\tau^A = \tau_p^A$ and $\tau^B \leq \tau_p^A$. With $\tau^A = \tau_p^A$ and $\tau^B \leq \tau_p^A$, (A3) and (A4) then imply $P^B(\tau^A, \tau^B, \tau^t) = P^A(\tau^A)$. On the other hand, the lowest value of $P^A(\tau^A, \tau^B, \tau^t)$ is achieved at $\tau^A = \tau^A$, $\tau^B = \tau_p^B$ and $\tau^t = \tau_p^B$, and $\tau^B = \tau^A$, and unless C is sufficiently large relative to B we must then have $P^B(\tau^A, \tau^B, \tau^t) = P^A(\tau^A)$, which is guaranteed if C is sufficiently small, then achieving efficient politically optimal tariffs in the WTO-Contract Game can be accomplished without the utilization of B’s non-violation right (i.e., we may then have $\tau^B = \tau_p^B$, but in this case negotiations between A and C must conform to reciprocity (i.e., the movement from $\tau_p^B$ to $\tau^A$ and from $\tau^A$ to $\tau^B$ must leave the terms of trade unaltered). In this way, Proposition 4 suggests that non-violation nullification-or-impairment rights can help to guard against backward stealing in an MFN environment by inducing later negotiations to abide by the reciprocity norm, and in doing so to preserve the value of earlier concessions. We note also that the requirement in Proposition 4 that

$$ C \text{ is not large relative to } B \text{ is suggestive of an efficiency-enhancing role for the “chief” – or “principal” – supplier rule utilized by the United States under the RTAA (see Tasca, 1938, pp. 146–147) and adopted as well in the GATT/WTO, whereby the largest suppliers to a market are typically granted the position of the early negotiating partners.}

More broadly, in light of this discussion it is evident that the backward-stealing and forward-manipulation problems which prevent governments from achieving efficient bargaining outcomes under sequential bilateral negotiations in MFN environments (Propositions 1 and 2) can in principle be addressed with the inclusion of features that have representation in the bargaining environment shaped by GATT/WTO rules. In particular, opportunities for renegotiation can in principle prevent the inefficiencies that arise as a result of the forward-manipulation problem (Proposition 3), while non-violation nullification-or-impairment rights operating in the presence of a reciprocity norm can in principle prevent the inefficiencies associated with backward stealing (Proposition 4).

7. Conclusion

Motivated by the structure of WTO negotiations, we analyze a bargaining environment in which negotiations proceed bilaterally and sequentially under the MFN principle. Our analysis proceeds in two steps. In a first step, we identify backward-stealing and forward-manipulation problems that arise when governments bargain under the MFN principle in a sequential fashion. We show that these problems impede governments from achieving the multilateral efficiency frontier unless further rules of negotiation are imposed. In our second step, we identify the WTO reciprocity norm and its nullification-or-impairment and renegotiation provisions as rules that are capable of providing solutions to these problems. In this way, we suggest that WTO rules can facilitate the negotiation of efficient multilateral trade agreements in a world in which the addition of new and economically significant countries to the world trading system is ongoing.

We have shown that the backward-stealing and forward-manipulation problems arise under very general circumstances, and that these problems can be interpreted as reflecting underlying incentives to manipulate the terms of trade. We reiterate here, though, that these problems can equally well be given an interpretation in terms of market access: each problem reflects the incentives of negotiating partners to position the balance of market access rights and obligations in a way that is disadvantageous for unrepresented governments. When interpreted from this perspective, the backward-stealing and forward-manipulation problems take on heightened practical relevance, because the balance of market access rights and obligations is a dominant theme in GATT/WTO discussions. And from this perspective, the potential importance of the role played by the WTO reciprocity norm and its nullification-or-impairment and renegotiation provisions in facilitating efficient bargaining outcomes may be appreciated.

Our findings are also consistent with the historical experience which led to the RTAA on which many of the essential features of the GATT/WTO are founded. As we have described, by 1934 the United States was frustrated with its experience in multilateral tariff bargaining, but also well-aware that the effectiveness of bilateral tariff bargaining could be undone by the twin problems of backward stealing and forward manipulation. Our results indicate that the RTAA and later the GATT and WTO succeeded in devising a method of bilateral tariff bargaining that was not prone to these fundamental problems. This is not to say that bilateral tariff bargaining under the GATT/WTO is necessarily free from problems such as foot-dragging and free-riding. In fact, Limao (2007) presents evidence consistent with the occurrence of foot-dragging in GATT tariff bargaining outcomes, and Ludema and Mayda (2009) report evidence consistent with free-riding in GATT market access negotiations (though Bagwell...
and Staiger (2010b) find little evidence of free-riding in the context of WTO accession negotiations). Rather, our results simply suggest that these problems would likely be far more wide-spread in the absence of the GATT/WTO rules highlighted by our analysis.

Finally, we have focused on the possibility of achieving efficient bargaining outcomes in various negotiating environments, but we have not characterized equilibrium outcomes in the environments where efficiency cannot be achieved. Hence our results do not indicate the severity of the inefficiency that arises when backward-stealing and forward-manipulation problems are present. We leave this to future research.

Acknowledgments

This paper has benefited from the detailed comments of two referees. We also thank Henrik Honn, Nuno Limao, Alberto Martin, Petros Mavroidis, Jim Rauch, Lex Zhao and seminar participants at Colorado, Columbia, Duke, Maryland, Oregon, Princeton, Stanford, Wisconsin, the NBER Summer Institute, the WTO and the Conference of the Japan Society of International Economics for very helpful comments, the European University Institute for their hospitality during which part of this paper was drafted, and the NSF for providing funding. All errors are ours.

Appendix A

In this Appendix, we provide proofs of all Lemmas and Propositions not established in the text.

Lemma 1. Any welfare triple can be achieved with an arbitrary market-clearing world price or with any instrument set at an arbitrary level.

Proof. Consider an arbitrary set of policies \((\tau, t)\) and associated welfare levels \(W^d(P^d(\tau, t), P^w, t)\) for \(j = (A, B, C)\) and market-clearing world price \(P^w(\tau)\). According to Eqs. (2) and (2a), a small change \(\Delta P^w\) in the market-clearing world price can be engineered as follows: (i) define the change in \(\tau\) for \(j = (A, B, C)\) according to \(\partial P^w(\tau, P^w, t)/\partial P^w \equiv 0\); and (ii) define the change in \(t\) for \(j = (A, B, C)\) according to \(\partial P^w(\tau, P^w, t)/\partial P^w \equiv 0\) implying \(\Delta t = -\partial \Delta P^w/\partial \tau\) by Eq. (2a), which by Eq. (2) then implies \(\Delta t = -M\partial \Delta P^w/\partial \tau\) and therefore implies as well \(\Delta P^w(\tau, \tau')\partial P^w(\tau, P^w, t)/\partial P^w = 0\) by Eq. (2a). Hence, the market-clearing condition (2) continues to be satisfied when these policy changes are made and the market-world price changes by \(\Delta P^w\). But by Eq. (3), these policy changes leave \(\Delta W^d = 0\) for \(j = (A, B, C)\). Next observe that the same changes in \(\tau\) for \(j = (A, B, C)\) and \(t\) for \(j = (B, C)\) described just above for engineering a small change in \(P^w\) can achieve the desired change in any particular tariff \(\tau\) for \(j = (A, B, C)\), since changing \(\tau\) for \(j = (A, B, C)\) according to \(\partial P^w(\tau, P^w, t)/\partial P^w \equiv 0\) implies \(\partial \Delta t/\partial P^w = -\partial \Delta P^w/\partial \tau\) and \(\partial \Delta t/\partial P^w = 1/\partial P^w = 1/\partial P^w\) for \(j = (B, C)\). And the ability to engineer a desired change in a particular transfer \(t\) for \(j = (B, C)\) is implied directly by the changes described just above, since according to those changes \(\partial \Delta t/\partial P^w = -\partial \Delta P^w/\partial \tau\).

Lemma 2. At each efficient point, the following restrictions apply:

\[(i)\; dW^d/\partial \tau > 0, j \in \{A, B, C\};
(ii)\; dW^d/\partial t > 0, \Delta W^d/\partial t = 0, j \in \{B, C\};
(iii)\; dW^d/\partial t > 0, j, \tau \in \{B, C\}.\]

Proof. (A1) states (R1)(i) hold. For (R1)(iii), note that (A1), (A2) and Eq. (4) imply \(dW^d/\partial t = 0\) for \(j = \{B, C\}\). But \(dW^d/\partial \tau = \sqrt{[1/2\partial t]W^d + \sqrt{\partial W^d}}\) \((\partial P^w/\partial \tau^d)\) and \(dW^d/\partial t = \sqrt{[1/2\partial t]W^d + \sqrt{\partial W^d}}\) \((\partial P^w/\partial \tau^d)\) for \(j = \{B, C\}\), and so \(\partial P^w/\partial \tau > 0, \partial W^d/\partial \tau^d > 0\) for \(j = \{B, C\}\). (R1)(iii) follows. Finally, (Eq. (4), (A1) and (R1)(iii) imply \(dW^d/\partial \tau^d < 0\) for \(j = \{B, C\}\), which gives (R1)(i).
Proposition 3. Under (A5), forward manipulation does not impede the attainment of the efficiency frontier in the Contract Renegotiation Game.

Proof. We first observe that, under (A5), A can do no better for itself in the Contract Renegotiation Game than to efficiently deliver $\vec{W}^{RD}$ to B and $\vec{W}^{CD}$ to C, and so it will choose to do so when this is feasible. We focus, then, on characterizing these feasibility conditions. That is, what conditions in addition to (A5) assure that there exists a stage-1 proposal $\{\vec{p}^A,\vec{p}^B,\vec{p}^C\}$ that would induce in the Contract Renegotiation Game a point on the efficiency frontier $\{\vec{W}^{RD},\vec{W}^{CD},\vec{W}^{CE}\}$ for which $\vec{W}^{HE} = \vec{W}^{RD}$ and $\vec{W}^{CE} = \vec{W}^{CD}$? By Lemma 1, for $\vec{p}^B$ fixed arbitrarily, there exists $\{\vec{p}^A(\vec{p}^B),\vec{p}^C(\vec{p}^B)\}$ that achieves $\{\vec{W}^{HE},\vec{W}^{HE},\vec{W}^{HE}\}$. We will say that $\vec{p}^B$ is consistent with (A1) and (A2) if (A1) and (A2) are satisfied at $\{\vec{p}^A(\vec{p}^B),\vec{p}^C(\vec{p}^B)\}$. Then under (A5), the SGPE of the Contract Renegotiation Game must be efficient if there exists a $\vec{p}^B$ consistent with (A1) and (A2) and a $\vec{p}^E \geq \vec{p}^A(\vec{p}^B)$ such that $\vec{w}^C(\vec{p}^E) = \vec{w}^C(\vec{p}^B)$ would induce the stage-2 proposal $\{\vec{p}^A(\vec{p}^E),\vec{p}^C(\vec{p}^E)\}$ and therefore lead to the payoffs $\{\vec{W}^{HE},\vec{W}^{HE},\vec{W}^{CE}\}$, and A can do no better than this. But the required conditions (other than (A5)) are simply those needed to satisfy the security constraint at the efficient point. □

References
