Abstract

International trade and tropical deforestation interact in multiple ways. This paper first presents a dynamic game whereby the South (S) exploits a resource in order to export. Because of negative externalities, the North benefits from trade liberalization only if the remaining stock is, in any case, diminished. Anticipating this, S exploits faster.

If tariffs can be contingent on the size of the remaining stock, however, the equilibrium agreement will reduce the deforestation rate. Combined with export subsidies, all forests can be conserved. Even though the parties cannot commit to future policies, they obtain the same outcome as if they could.

Key words: Exhaustible resources, deforestation, dynamic games, autarky, trade agreements, environmental conservation, renegotiation.

JEL: F18, F13, F55, Q56, Q37.

Email: harstad@stanford.edu. I have benefited from many discussions and suggestions, and I am especially grateful for the detailed comments from Arild Angelsen, Geir Asheim, Kyle Bagwell, Torfinn Harding, Henrik Horn, Giovanni Maggi, Rick van der Ploeg, and Bob Staiger. Kristen Vamsæter and Valer-Olimpiu Suteu provided excellent research assistance and Frank Azevedo helped with the editing. The research received funding from the European Research Council under the European Union’s Horizon 2020 research and innovation program (grant agreement 683031).
I. Introduction

In 2019, Brazil led the Mercosur trade bloc to conclude its largest trade agreement ever with the European Union. While the trade negotiations concluded, deforestation rates increased and the forest fires gained international media attention. Deforestation in the Brazilian Amazon has increased every year since that time, even though the agreement hasn’t even been ratified. Fearing deforestation, critics argue that the treaty with the EU should not be ratified in their current form. On June 5, 2023, Reuter reported that France seeks "additional commitments, notably on respecting EU rules on deforestation, before it can back it."\(^1\)

Trade can lead to deforestation,\(^2\) and deforested land can lead to more trade. Analysing the two-way interaction requires a dynamic model. My model includes the South (S), endowed with a depletable resource, and the North (N), preferring conservation. The resource can be oil or coal but, to fix ideas, I will refer to tropical forests. Trade liberalization increases S’s prices for timber and beef, so exploitation increases when a free trade agreement (FTA) is signed. For that reason, an FTA is socially optimal only if the remaining resource stock has already diminished, so that the additional damage to N, when the remaining resource is depleted, is less than the gains from trade. In other words: Trade causes depletion, and depletion leads to an FTA.

Anticipating the FTA that will be signed when the resource stock is diminished, S faces an additional incentive to exploit, even before the FTA is signed. This incentive, in turn, can persuade N to sign the agreement right away. If N could commit, N would commit to autarky when the remaining stock is large. But because S understands that N will always allow for trade later, if the stock is diminished, S is willing to exploit, and therefore, N is willing to trade. The equilibrium is that N and S sign the FTA, and the resource is depleted, even when the damages are larger than the gains from trade.

Even in this dynamic game, all negative findings are reversed under a contingent trade agreement (CTA). The CTA allows the parties to negotiate tariffs, and thus the allocation of gains from trade, that depend on the resource stock. In equilibrium, S will face no tariff when the stock is large while N will face no tariff when the stock is small. Even when the tariff functions are renegotiation proof and thus credible, this difference motivates S to reduce the deforestation rate. If the tariffs can be accompanied by N’s export subsidies, then there is no limit to how large the effective transfer from N to S can be, and then the first best can be implemented. The efficiency result holds regardless of the allocation of bargaining power, even though that allocation influences how the gains from trade are split.

The Online Appendix shows that the equilibrium CTA implements the same outcome that N and S would have obtained if had they been able to commit to a trade agreement that was conditional on the resource stock. Consequently, the optimal CTA is not an arbitrary design from which N and S can make further improvements: It implements the first best if export subsidies are available, and the second best

\(^1\)https://www.reuters.com/world/eu-commission-says-mercosur-eu-deal-priority-2023-06-05/

\(^2\)This is empirically supported by, e.g., Faria et al. (2016), Pendrill et al. (2019), and Abman and Lundberg (2020).
otherwise. The Online Appendix also shows that the CTA can be robust to unilateral requests to renege, and it explains why exhaustability is the key assumption.

The structure of CTAs is related to a 2020 policy initiative from France and Netherlands: In a recent "non-paper," they recommend that the implementation of trade agreements should proceed step-wise and hinge on the gradual implementation of sustainability requirements. The non-paper concludes by arguing that the WTO should take on a special role. The analysis below suggests that more forests can be conserved if the WTO relaxes this prohibition for conservation purposes.

When trade causes environmental problems, other scholars have recommended trade sanctions (Barrett, 1997), border tax adjustments (Ludema and Wooton, 1994), and climate clubs (Nordhaus, 2015). Copeland et al. (2022) explain, however, that the traditional solutions are neither ex post efficient, nor credible if the resource is depletable: after depletion, everyone benefits from trade. Hsiao (2022) quantifies the inefficiencies.

Some studies of issue linkages rely on double punishments that may not be renegotiation proof (Edel-ington, 2002; Limão, 2005; Maggi, 2016). The CTA, in contrast, exploits the fact that there are multiple Pareto optimal allocations of trade, and the selected allocation can be a function of the history. This method is inspired by how cooperation is implemented in dynamic games when the parties can renegotiate. To make punishments credible, one may need to require that the punishment payoffs continue to be on the Pareto frontier, although the payoff must be unattractive for the party that has defected (Mailath and Samuelson, 2006). I combine this logic with the theory of issue linkages because, in my analysis, the Pareto frontier refers to various allocations of the gains from trade, while defection refers to resource depletion. The problem is nontrivial because the (ex post) Pareto frontier (i.e., the gains from trade) expands if the resource is depleted.

A companion paper (Harstad, 2023) verifies that the CTA can motivate conservation also in a standard trade model with concave utility functions and distortionary tariffs. That model allows for many products, countries, and collaborators, permitting a quantitative analysis. That analysis is static, however, and thus it captures neither that S may want to deplete to obtain an FTA, nor how the CTA can be designed so as to postpone deforestation to later periods.

II. The Model

The South. Let S be a country endowed with a resource stock that can be depleted over time. At the beginning of time \( t \in \{1, 2, \ldots \} \), the stock is \( R_t \), the part that has been exploited is \( X_t \), so \( R_0 = R_t + X_t \) is the original size. When S exploits \( x_t \in [0, R_t] \),

\[
R_{t+1} = R_t - x_t \quad \text{and} \quad X_{t+1} = X_t + x_t.
\]

---

3 https://nl.ambafrance.org/Non-paper-from-the-Netherlands-and-France-on-trade-social-economic-effects-and-

4 Mailath and Samuelson (2006:122) write: "should [the players] ever find themselves facing an inefficient continuation equilibrium, whether on or off the equilibrium path, they can renegotiate..." (p. 122). Regarding the solution to this problem, "the key...is to select punishments that reward the player doing the punishing" (p. 135).
The resource can represent oil or coal but, to fix ideas, I refer to $R_t$ as the remainder of the rainforest, $X_t$ as the land that has already been logged and converted to agriculture, and the timber currently logged is proportional to $x_t$. For simplicity, S’s agricultural produce (beef) equals its amount of converted land, $X_t$. In autarky, $\alpha \geq 0$ represents the (present-discounted) agricultural value of land, if $(1 - \delta) \alpha$ measures S’s per-period utility of the food produced per unit of land and $\delta \in (0, 1)$ is the discount factor. In addition, $\beta \geq 0$ is S’s marginal benefit of the extracted units (timber), while $c$ is the marginal cost of exploitation. The cost $c$ may include the physical as well as the (present-discounted value of the) environmental cost to S when $R_t$ is reduced by a unit.

In autarky, S is a single decision maker maximizing its continuation value:

$$V_S^{AUT}(R_t) \equiv \max_{x_t^{AUT} \in [0, R_t]} (1 - \delta) (X_t + x_t) \alpha + \beta x_t - cx_t + \delta V_S^{AUT}(R_{t+1}).$$

The linearity in $x_t$ implies that the autarky choices are simple:

$$x_t^{AUT}(R_t) = \begin{cases} 0 & \text{if } \alpha + \beta \leq c \\ X_t \in [0, R_t] & \text{if } \alpha + \beta = c \\ R_t & \text{if } \alpha + \beta > c \end{cases}. \tag{1}$$

The North. N is S’s potential trading partner. Just like S, N can experience costs and benefits from the exploitation. In particular, N faces damage $d > 0$ for each unit that S logs. Equivalently, $d$ represents N’s marginal present-discounted value if a resource unit is conserved forever.

N’s marginal value from beef is $(1 - \delta) \pi$. That is, $\pi$ is N’s present-discounted value of consuming a unit of S’s agricultural products in every future period. N’s marginal benefit from the extracted resource (timber) is $\beta$.

If $\pi \geq \alpha$, it is socially optimal that the beef ($X_t$) be exported to N, and if $\beta \geq \beta$, it is socially optimal that the timber ($x_t$) be exported to N. Both inequalities are assumed to hold weakly.\(^5\)

I assume that the seller sets the price. Thus, with a free trade agreement (FTA), S receives $(1 - \delta) \pi$ for each exported unit of beef in every period, and $\beta$ for each unit of timber. To S, exploitation for trade is strictly beneficial if and only if (iff) $\pi + \beta > c$. Exploitation is socially inefficient when we assume

$$\pi + \beta < c + d. \tag{2}$$

N benefits $e > 0$ from getting access to S’s market. We can endogenize $e$ as follows. Suppose that N has the capacity to export $\psi \geq 0$ units of machines in every period to S. If the citizens of S are willing to pay $\omega \geq 0$, and N’s marginal production cost is $\kappa \in [0, \omega]$, then N charges $\omega$ and captures the export’s entire present-discounted value $(\omega - \kappa) \psi / (1 - \delta)$, henceforth defined as $e$.

We have a general-equilibrium model, and trade is balanced, if we introduce a numeraire good that can be used as a currency. For example, countries may trade cookies or labor services. For each country, the value parameters above are measured relative to the country’s value of the numeraire.

\(^5\)This assumption is without loss of generality for the present analysis because if, for example, $\pi < \alpha$, S’s beef will not be exported and the realized gains from trade will be zero, i.e., the same as when $\pi = \alpha$. 

First Best. In every period $t$, the gains from trade are given by

$$(1 - \delta) c + (1 - \delta) (\pi - a) (X_t + x_t) + (b - \bar{b}) x_t > 0.$$  

The first-best outcome is simply that the parties trade and that S conserves in every period (i.e., $x_t = 0$).

Timing. In each period, $t$, the parties first bargain whether to open up for trade, if they haven’t opened up already. Second, S decides on $x_t \in [0, R_t]$. S is assumed to conserve whenever indifferent. Finally, trade and consumption take place.

Bargaining Solution. Both countries must agree to open up for trade, but side transfers can be used if the countries agree on an FTA. Let $\alpha \in [0, 1]$ measure S’s share of the bargaining surplus. Section III.E explains how tariffs can substitute for transfers, and how the results survive without transfers.

Equilibrium. S’s strategy is mapping from the set of histories to $x_t \in [0, R_t]$. N does not take any action: N must agree to trade, but N and S are simply sharing the gains from liberalization if the gains are positive. The below inefficiency results (Propositions 1–3) hold for all subgame-perfect equilibria (SPEs). The efficiency result in Section IV holds despite the restriction to Markov-perfect equilibria (MPEs). It is, at that stage, natural to focus on MPEs given the importance of the state variable $R_t$.

III. Free Trade Agreements

Before characterizing the equilibrium in detail, it is useful to start with three observations.

III.A. Trade Causes Exploitation

With free trade forever, S faces an alternative to consuming the resource domestically. In autarky, S exploits $x_{AUT}^t > 0$ only if $a + b > c$. With trade, S exploits also if $\pi + \bar{b} > c$. It follows that S exploits more with trade than in autarky.

**Proposition 1.** Moving from autarky to free trade causes depletion:

$$x_{FT}^t = R_t \geq x_{AUT}^t = 0, \text{ if } c \in [a + b, \pi + \bar{b}).$$  

If $\pi + \bar{b} \leq c$, then $x_{AUT}^t = x_{FT}^t = 0$. If $a + b > c$, then $x_{AUT}^t = x_{FT}^t = R_t$.

All proofs are in the Appendix.

Proposition 1 implies that no SPE can implement the first best with an FTA and no exploitation.

It is easy to see that the equilibrium survives also if trade liberalization is reversible and must be decided on in every period. As soon as the parties trade in one period, S exploits by choosing $x_t = R_t$. Thereafter, when the resource is depleted, trade is unambiguously efficient in every period. A threat to not trade after depletion is not credible.\(^6\)

\(^6\)If the transfers/tariffs can vary, there exist SPEs in which N pays S in every period as long as S conserves, if just $\delta$ is sufficiently large. Such SPEs cease to exist if the transfer, as here, cannot be conditioned on S’s action.
III.B. Exploitation Causes Trade

Since utilities are transferable, it is the sum of the two continuation values that determines whether an agreement is beneficial. By comparing the autarky payoffs following (1) and the FTA payoffs following Proposition 1, we can conclude that the benefit from an FTA can be negative if \( R_t \) is large. When (3) holds, the resource will be depleted with an FTA but not in autarky.

If \( R_t \) is already diminished, however, the additional damage is small and outweighed by the gains from trade. Thus, there exists a threshold, \( R^* \), so that the FTA is socially valuable if \( R_t \leq R^* \).

**Proposition 2.** Suppose trade influences \( x_t \) (i.e., (3) holds). The social value of the FTA at time \( t \), relative to autarky, decreases in \( R_t \) and is positive if:

\[
R_t \leq R^* = \frac{c + (\pi - a)R_0}{c + d - a - b}
\]  

(4)

The proposition describes a second-best outcome: Given the inefficiency uncovered by Proposition 1, it is socially optimal with trade iff the resource has already been exploited so much that the remainder \( R_t \) is small and inequality (4) holds. In this case, the parties strictly benefit from trade, despite the fact that trade will motivate further exploitation.

III.C. Exploit to Trade

The parties cannot commit. If it should happen that the resource is exploited anyway, so that (4) holds, N and S will find it optimal to trade.

S anticipates that when (4) holds, it will be able to trade. Even if \( a + b < c \), so that S finds it too costly to exploit in autarky, this cost is worth paying if \( R_t \) is already small or if the gains from trade are large.

**Proposition 3.** S benefits from exploiting to obtain an FTA if the gains from trade are large or \( R_t \) is small, i.e., if:

\[
R_t < \hat{R} \equiv \delta \alpha e + (\pi - a)R_0 \frac{c - b - a}{c - b - a} \text{ or } a + b > c.
\]  

(5)

By comparison,

\[
\hat{R} > R^* \iff d > \hat{d} \equiv \frac{1}{\delta \alpha - 1} \left( c - b - a \right) + \tilde{b} - \tilde{b}.
\]

When \( R_t \in (R^*, \hat{R}) \), the second-best outcome, characterized by Proposition 2, cannot be sustained by any SPE. Even if the FTA is not socially valuable, S can always obtain a larger continuation than in autarky by first exploiting the resource and then trade. This situation is more likely to occur if S’s bargaining power (\( \alpha \)) is large.

Proposition 3 states that S is "willing" to exploit but, on the equilibrium path, N and S find it optimal to sign an FTA as soon as N expects that S will, in any case, exploit the resource.
III.D. Equilibrium

The propositions above hold for every SPE. Proposition 1 states that the first best cannot be supported as an SPE. Proposition 2 describes the second best, given the inefficiency uncovered by Proposition 1, but Proposition 3 states that not even the second best can be supported as an SPE: There is no SPE without trade or exploitation if $R_t \in \left(R^*, \hat{R}\right)$. There are multiple SPEs, but all lead to trade and exploitation.\(^7\)

To offer a sharp characterization of the outcome, I henceforth characterize the MPE in pure and linear strategies. In the MPE, N and S sign the FTA iff either (4) or (5) holds (or both), i.e., in region A, B, or C in Fig. 1.

The MPE also specifies S’s action off the equilibrium path, that is, the threat point, or the default extraction level, $x_D$, that S will choose at the extraction stage if the parties fail to sign an agreement.

**Proposition 4.** In equilibrium, the FTA is signed, and S exploits, iff the gains from trade are large or $R_t$ is small:

if $d \leq \hat{d}$, the condition is $\frac{e + (\overline{\alpha} - a)R_0}{R_t} > c + d - b - a$, and $x_D = 0$,

if $d > \hat{d}$, the condition is $\frac{e + (\overline{\alpha} - a)R_0}{R_t} > \frac{c - b - a}{\phi \alpha}$, and $x_D = R_t \cdot \min\{\phi, 1\}$,

where $\phi = \frac{d - \hat{d}}{c + d - b - a + \frac{b - a}{1 - \phi}}$. \(^6\)

In region A in Fig. 1, $d < \hat{d}$, and S is not willing to exploit in order to obtain an FTA when (4) fails. In this situation, N and S trade iff $R_t \leq R^*$, and S’s default extraction level is $x_D = 0$. This situation corresponds to the first case in the proposition.

In the second case, and in region B, N and S trade even if the FTA is socially suboptimal, because S will otherwise exploit in order to obtain an FTA later (Proposition 3). The threat point is that S exploits the fraction $\phi > 0$. The fraction may be less than 1 because if $x_D$ were very large, the bargaining surplus would be large, and – anticipating the fraction $\alpha$ of it – S would be willing to conserve. If $\phi < 1$, S stays indifferent, in equilibrium.\(^8\)

In region C, $R_t$ is even smaller, and both (4) and (5) hold. In this case, S would be willing to exploit $R_t$ in order to trade, but that is not necessary. After all, N and S jointly benefit from the FTA, even without that threat. In fact, N and S would negotiate an FTA and S would obtain the fraction $\alpha$ of the total surplus even if the default outcome were that S would not exploit. That surplus, it turns out, can be less than what S can obtain from first exploiting and then negotiating an FTA. The reason for why it can be less is that if the default outcome were $x_D = 0$, S would have to compensate N for the damages N would face given that the FTA will cause exploitation. If S, instead, exploits first, the damage is sunk

\(^7\)For example, there are SPEs in which N and S agree to trade at some future time, or in every $\Delta t$ period, where $\Delta t$ is so small that S prefers to conserve when there are fewer than $\Delta t$ periods left.

\(^8\)In the mixed-strategy equilibria, S’s expected default extraction level is $x_D$, described by Proposition 4.
Figure 1: An FTA is socially optimal if $R_t < R^*$, but S is willing to exploit to obtain one if $R_t < \hat{R}$. In region B, N and S trade even if the total surplus is larger in autarky.

and no such compensation can be requested. The latter option is preferable to S when $d > \hat{d}$. So, in region C, S is willing to exploit in order to obtain a better FTA.

III.E. Transfers and Tariffs

The results above continue to hold, qualitatively, even if N and S cannot use transfers at the negotiation stage. For example, Proposition 1 remains unchanged, the inequality in Proposition 2 simplifies to $e/R_t \geq d$, and (5) in Proposition 3 is replaced by $(\pi - a) R_0/R_t > c - a - b$.

Here, transfers can take the form of tariff adjustments. If $\tau_N \leq 1$ measures S’s ad valorem import tariff on N’s export, then consumers in S are willing to pay only the fraction $(1 - \tau_N)$ for N’s goods, relative to how much they would have paid without any tariff. N will find it necessary (and optimal) to reduce the price by this fraction and, therefore, N loses and S gains.

A tariff ($\tau_S$) in N on S’s beef is similarly improving N’s, but worsening S’s, terms of trade.

If there are no export subsidies, S is willing to export only if $(1 - \tau_S) \bar{\pi} \geq a$. (There will be a similar constraint on $\tau_N$.) These constraints can be ignored if the tariffs can be accompanied by export subsidies (Grossman and Helpman, 1995): If $s_S$ is an ad valorem export subsidy in S, S’s producers are willing to export as long as $(1 + s_S)(1 - \tau_S) \bar{\pi} \geq a$. Alternatively, the transfers can be arbitrarily large if, instead of export subsidies, we permit import subsidies.

With both tariffs, the payoffs after signing the agreement become:

$$V^{FTA}_S(R_t, \tau) = \tau_N e + (1 - \tau_S) \bar{\pi} R_0 + (\bar{b} - c) R_t = V^{FTA}_S(R_t, 0) + \tau,$$ and

$$V^{FTA}_N(R_t, \tau) = (1 - \tau_N) e + \tau_S \bar{\pi} R_0 - d R_t = V^{FTA}_N(R_t, 0) - \tau,$$ where

$$\tau \equiv \tau_N e - \tau_S \bar{\pi} R_0.$$
Note that \( \tau \) essentially represents a transfer from N to S.\(^9\)

In the dynamic game, the gains from trade include the gains from starting with the FTA at time \( t \) instead of at \( t+1 \). In the meanwhile, S exploits \( x^D \). The equilibrium level of \( x^D \) is thus going to influence the negotiated transfer. This is evident in the next proposition, which presents the equilibrium transfer implemented by the tariffs.

**Proposition 5.** If N and S sign an FTA at time T, then \( \tau_N \) is smaller or \( \tau_S \) is larger if \( R_T \) is large:

\[
\tau(R_T) = \alpha c - (1 - \alpha) \Delta_a R_0 - R_T \cdot \left\{ \begin{array}{ll}
\alpha d - (1 - \alpha) (c - a - \bar{b}) & \text{if } \phi < 0 \\
\frac{\bar{b} - b + [c - b - a(1/\delta - 1)]}{(\bar{b} - b)(1 - \alpha)} & \text{if } \phi \in [0, 1] \\
\frac{c + b - a(1/\delta - 1)}{(\bar{b} - b)(1 - \alpha)} & \text{if } \phi > 1
\end{array}\right.
\]

Intuitively, when N and S negotiate whether to sign the FTA, the equilibrium transfer from N to S will reflect the bargaining strength \((\alpha)\), the gains from trade, and the payoffs in the outside option (i.e., in autarky).

The larger \( R_T \) is, the larger is S’s payoff from the FTA, but the smaller N’s payoff is from the FTA when the FTA causes exploitation (i.e., when (3) holds). The equilibrium transfers or tariffs ensure that the parties will obtain comparable gains from the bilateral deal. Consequently, \( \tau_N \) must decrease in \( R_T \), or \( \tau_S \) must increase in \( R_T \) (or both).

**IV. Contingent Trade Agreements**

**IV.A. Feasibility**

With tariffs, gains from trade can be distributed in alternative ways. Once the parties have agreed to trade, and the gains are allocated according to \((\tau_N, \tau_S)\), every such allocation is renegotiation proof in the following sense: any change in \( \tau \) will harm and thus be vetoed by (at least) one of the parties.

However, we may not want to impose the restriction that the equilibrium allocation of gains, or the pair \((\tau_N, \tau_S)\), must be constant. This section permits the parties to negotiate tariffs that are functions not only of \( R_T \) at the time, \( T \), when they negotiate, but also of every smaller \( R_t \in [0, R_T] \) that is imaginable for future dates (even off the equilibrium path).

**Definition:** A CTA, negotiated at time \( T \), specifies tariffs \( \tau_S (R_t; R_T) \) and \( \tau_N (R_t; R_T) \), that can depend on the current \( R_t \) as well as on \( R_T \), unless the parties agree on different tariffs.

For every \( R_t \), the agreement must give S at least the same payoff as S can obtain in autarky. In addition, the tariff functions must be renegotiation proof. This requirement rules out, for example, a punishment strategy in which S will no longer be able to export if \( R_t \) has been reduced. As observed above, if \( R_t = 0 \), it is always (ex post) better for N and S to trade. The agreement is renegotiation proof if the equilibrium payoff pair is on the Pareto frontier for every \( R_t \) that is feasible at \( t > T \).

\(^9\)To facilitate this transfer, there is no need for introducing tariffs on timber \((x_t)\) also.
A challenge is that, even if there is no exploitation after the agreement is signed, the sum of the gains from trade, \( e + (\pi - q)X_t \), increases in \( X_t \). Thus, the more that has been depleted, the larger are the gains from trade that can be shared. If S receives a constant fraction of this cake, S would face a strong incentive to exploit.

Even if the total gain increases, the gain allocated to S can decrease. This decrease is possible if S’s tariff is a function that declines in the stock, while N’s tariff is a function that increases in the stock. If \( \tau \) increases sufficiently fast in \( R_t \), then S has an incentive to conserve rather than to exploit.

**Lemma 1.** S is willing to conserve s.t. \( x_t = 0 \) \( \forall t \geq T, R_t \leq R_T \), iff:

\[
\frac{\partial \tau(R_t; R_T)}{\partial R_t} \geq \pi - \frac{b}{\alpha} - c \forall R_t \leq R_T, \text{ where } \tau(R_t; R_T) = \tau_N(R_t; R_T) e - \tau_S(R_t; R_T) \pi X_t.
\]

Thus, the CTA can specify a rule that allocates the gains from trade to S as long as S conserves, and to N if S exploits its resource. When (8) holds, this reallocation of the gains occurs so fast when \( R_t \) is reduced that S is better off conserving than depleting the resource.

**IV.B. The First Best with Export Subsidies**

The lemma states that it is possible to design an agreement that motivates conservation. The next result states that N and S will indeed sign such an agreement in equilibrium, as long as the tariffs are permitted to be a function of the remaining stock, \( R_t \). The intuition for this statement is simply that conservation is socially efficient, and thus both N and S can benefit from an agreement that motivates conservation when the parties can use side transfers (e.g., tariffs). N and S will share the total surplus according to their respective bargaining strengths.

The following proposition describes the unique MPE (\( \tau \) and \( x^D \)) in pure strategies. The tariffs are written as a function of the stock that exists at the time of negotiations, \( R_T \). If future stocks are different, the tariffs will also change in line with (8). Because (8) can be respected by a continuum of functions, the proposition does not specify exactly how steeply the tariffs will change if (off the equilibrium path) \( S \) extracted rather than conserved.

**Proposition 6.** Consider a subgame starting at time \( T \) without a CTA. In equilibrium, N and S sign a CTA and implement the first-best outcome with \( x_t = 0 \forall t \geq T \). The tariffs respect (8) and:

\[
\tau(R_T; R_T) = \begin{cases} 
\alpha e - (1 - \alpha)(\pi - a)R_0 + (1 - \alpha)(\pi - a)R_T & \text{and } x^D = 0 \quad \text{if } \varphi < 0 \\
\alpha e - (1 - \alpha)(\alpha - a)R_0 + \frac{\alpha b + c}{\alpha - a}R_T + (\alpha - a)R_T & \text{and } x^D = \varphi R_T \quad \text{if } \varphi \in (0, 1) \\
\alpha e - (1 - \alpha)(\pi - a)R_0 + [(1 - \alpha)(\pi + b - c) + \alpha d]R_T & \text{and } x^D = R_T \quad \text{if } \varphi > 1
\end{cases}
\]

where \( \varphi = \frac{\delta a (\pi - a) + a + b - c}{\delta a (\pi - a) + \delta a \frac{\alpha b^2 - \pi - b}{1 - \varphi}}. \)

Interestingly, \( \tau(R_T; R_T) \) increases in \( R_T \). The intuition is that if \( R_T \) is large, N’s benefit from an agreement that leads to conservation is also large. The larger benefit to N implies that, at the bargaining
stage, N will accept transfers to S, or will accept to face tariffs when exporting to S. Thus, $\tau(\cdot)$ increases in both arguments. The Appendix proves that $\tau(R_T; R_T)$ permits (8) to hold for every $R_t \in [0, R_T]$.

It is always first best to conserve the entire resource in the simple model studied here. Full conservation is feasible by letting S obtain a large share of the gains from trade when $R_t$ is large, but a smaller share when $R_t$ is small. If $R_T$ is very large, then the gain to S might need to be larger than the total gains from trade, requiring $\tau_N$ to be so large that N must subsidize its export for the producers to be willing to sell, or that $\tau_S < 0$, so that N subsidizes import. When $\tau_N$ is accompanied by an export subsidy in country N, or N subsidizes import, then there is no limit to how large $\tau$ can be, and there is no limit to how much one can conserve. N agrees to the large $\tau$, in equilibrium, because it motivates conservation.

**IV.C. Postponed Exploitation in the Second Best**

Export subsidies are rarely used in practice. They are generally prohibited by the WTO. When export and import subsidies cannot be used, the transfer from N to S is limited by the gains from trade. These gains limit how much S can be persuaded to conserve by simply being allocated the gains from trade. This limit is illustrated by the dotted line in Fig. 2.

**Proposition 7.** Consider a subgame at $T$ without a CTA. Suppose subsidies are not available and, for simplicity, that $\alpha = 0$. In equilibrium, N offers a CTA that S immediately accepts:

(i) The tariffs are in line with Proposition 6 and $x_t = 0$ for every $t \geq T$ if:

$$\frac{e + (\pi - a)R_0}{R_T} \geq \alpha + \frac{b - c}{\alpha + b - c} \Leftrightarrow$$

$$R_T \leq \tilde{R} \equiv \frac{e + (\pi - a)R_0}{\alpha + \frac{b - c}{\alpha + b - c}} \text{ or } \alpha + b < c.$$ 

(ii) When (9) fails, i.e., $R_T > \tilde{R} > 0$, then, for every $t \geq T$,

$$x_t = \left( R_t - \tilde{R} \right) \gamma, \text{ where } \gamma = \frac{\alpha + \frac{b - c}{\alpha + b - c}}{\alpha + b - c + \frac{b - c}{\alpha}} \in (0, 1),$$

and, on the equilibrium path $\tau_S = 0$ and $\tau_N = 1$.

Eq. (9) can be rewritten as:

$$\frac{e + (\pi - a)(R_0 - R_T)}{R_T} \geq a + \frac{b - c}{c}.$$ 

With this reformulation, part (i) in Proposition 7 states that if $a + b < c$, so that S does not want to reduce $R_t$ in autarky, the CTA can always ensure that there is no further exploitation, even if export subsidies cannot be used. If $a + b > c$, so that S would exploit in autarky, the CTA can still motivate conservation, so that $R_t \in (0, R_0) \forall t$, but the amount of resource conservation, $\tilde{R}$, is limited by the value of trade.

In Fig. 2, the solid decreasing line illustrates that the revenue from trade is larger when $R_t$ is small, but the dashed upward-sloping line illustrates how $V_S$ must vary with $R_t$ to motivate conservation. When $V_S$ is limited by the first line, conservation is limited to $\tilde{R}$.
Figure 2: When $V_S$ is limited by the gains from trade, which decreases in $R$ (solid line), at most $\bar{R}$ can be conserved. Because $V_S$ must increase in $R$ to motivate conservation, $V_N$ must decline in $R$ (dashed line). $N$’s payoff is, nevertheless, increasing in $R$ when the avoided damage is taken into account (upward-sloping solid line).

Part (ii) states that if $R_T > e > 0$, the gains from trade are insufficient to motivate full conservation. The CTA can nevertheless be used to motivate a slower extraction rate. The speed at which $S$ extracts can be reduced by allocating most of the gains from trade to $S$ as long as $S$ sticks to (11). If $S$ exploits faster, $S$ will face higher tariffs. The larger are the gains from trade, $b - \bar{a}$, the more it is possible to persuade $S$ to conserve in each period.

A similar result holds under the restriction $\tau_N = 0$, so that $N$ can, at best, offer $S$ free trade: Proposition 7 holds if just $e$ is replaced by zero.

Remark on $\alpha$. If $\alpha > 0$, the equilibrium $x_t$, when $R_T > \bar{R}$, would be larger than the $x_t$ described by (11). When $R_T > \bar{R}$, the level of $\alpha$ is thus not only affecting the distribution of surplus, but also efficiency: a larger $\alpha$ is less efficient because $S$, then, requests a CTA that tolerates faster extraction.

IV.D. Comparison to the FTA

Even though trade caused exploitation and exploitation caused trade in the dynamic model of Section III, the CTA overturns all the benchmark results.

Corollary 1. With a CTA, Propositions 1–5 are overturned:

1. The agreement leads to more conservation than in autarky.
2. The agreement is more valuable when $R_T$ is large.
3. $S$ conserves to maintain the most attractive agreement.
(4) In equilibrium, $S$ conserves when the gains from trade are large and $R_t$ is small.

(5) $\tau_N$ strictly increases or $\tau_S$ strictly decreases (or both) in $R_T$ as well as in $R_t$.

V. Conclusion

When free trade leads to resource exploitation, it is worthwhile only if the remaining stock is, in any case, small. That fact can motivate resource owners to reduce the stock to the point at which trade is acceptable for everyone. In the dynamic model, the parties trade and the resource is exploited even when autarky would be more efficient.

With a contingent trade agreement, in contrast, the negative results are reversed and exploitation can be postponed. This possibility should be kept in mind by scholars studying trade and environmental problems, but also by policymakers, public officials, and activists who struggle with how to balance trade and conservation. In fact, the analysis can shed light on how one might implement the proposal in the recent non-paper by France and Netherland, mentioned in the Introduction. The analysis also uncovers a possible role for the WTO, because by permitting export subsidies, more can be conserved.
References


Appendix: Proofs

Proposition 1.
When $R_t = 0$, a best response for both parties is to trade. When $S$ obtains the fraction $\alpha$ of the total gains from trade in addition to $S$’s autarky payoff, then:

$$
V^{FTA}_S(0) = V^{AUT}_S(0) + \alpha \left[ e + (\bar{\pi} - \bar{a}) R_0 \right] = \alpha R_0 + \alpha c + \alpha (\bar{\pi} - \bar{a}) R_0,
$$

because

$$
V^{AUT}_S(0) = \alpha R_0.
$$

When $R_t > 0$ and the parties trade, $S$ solves:

$$
V^{FTA}_S(R_t) = \max_{x^{FTA}_t \in [0, R_t]} (1 - \delta) (X_t + x_t) \bar{\pi} + \bar{b} x_t - cx_t + \delta V_S(R_{t+1}),
$$

which implies, regardless of whether $V_S(R_{t+1}) = V^{FTA}_S(R_{t+1})$ or $V_S(R_{t+1}) = V^{AUT}_S(R_{t+1})$, that

$$
x^{FTA}_t(R_t) = \begin{cases} 
0 & \text{if } \bar{\pi} + \bar{b} \leq c \\
x_t \in [0, R_t] & \text{if } \bar{\pi} + \bar{b} = c \\
R_t & \text{if } \bar{\pi} + \bar{b} > c
\end{cases}.
$$

Proposition 2.
When (3) holds, autarky leads to the total payoff $V^{AUT}(R_t) \equiv V^{AUT}_S(R_t) + V^{AUT}_N(R_t) = (R_0 - R_t) \alpha$, while the FTA leads to depletion and the total payoff

$$
V^{FTA}(R_t) \equiv V^{FTA}_S(R_t) + V^{FTA}_N(R_t) = e + \bar{\pi} R_0 + (\bar{b} - c - d) R_t,
$$

which is larger if:

$$
e + \bar{\pi} R_0 + (\bar{b} - c - d) R_t > (R_0 - R_t) \alpha \Leftrightarrow (4).$$

Proposition 3.
Without trade at $t$, $S$ is strictly better off with $x_t = R_t$ than with $x_t = 0$ and autarky forever if:

$$(1 - \delta) R_0 \alpha + (\bar{b} - c) R_t + \delta V^{FTA}_S(0) > \alpha (R_0 - R_t) \Leftrightarrow (5).$$

Proposition 4.
Trade equilibria. Consider, first, the situation in which $N$ and $S$ trade at every $R \leq R_t$.

Bargaining surplus. Let $x^D_t = \eta R_t$, with $\eta \in [0, 1]$, measure $S$’s extraction after disagreement. (In principle, $\eta$ can be a function of $R_t$.) The proof proceeds by deriving the fixed point where $S$’s best response, given $\eta$, coincides with $\eta$.

Given $\eta$, if the parties disagreed at $t$, but expect to agree at $t + 1$, the sum of disagreement payoffs is:

$$
V^{DIS}(R_t) = (1 - \delta) (X_t + \eta R_t) \alpha - (c + d - \bar{a}) \eta R_t + \delta V^{FTA}((1 - \eta) R_t)
$$

$$
= (1 - \delta) (R_0 - R_t) \alpha - [(1 - \delta) (c + d - \bar{a} - \bar{b}) + (\bar{b} - \bar{b})] \eta R_t + \delta [e + \bar{\pi} R_0 + (\bar{b} - c - d) R_t].
$$

The total gains from agreeing at $t$, (12), minus disagreement payoff $V^{DIS}(R_t)$, is:

$$
\Delta_R = (1 - \delta) [e + \bar{\pi} R_0 + (\bar{b} - c - d) R_t] - (1 - \delta) X_t \alpha + [(1 - \delta) (c + d - \bar{a} - \bar{b}) + (\bar{b} - \bar{b})] \eta R_t
$$

$$
= (1 - \delta) [e + (\bar{\pi} - \bar{a}) R_0 + (\bar{a} + \bar{b} - c - d) R_t] + [(1 - \delta) (c + d - \bar{a} - \bar{b}) + (\bar{b} - \bar{b})] \eta R_t.
$$

S’ best response. Consider, first, the case in which $x^D_t = 0$ is among $S$’s best responses. Even after disagreement at $t$, $N$ and $S$ will agree at $t + 1$ and, then, $S$ expects $S$’s default payoff plus fraction $\alpha$ of (13):

$$
V^{FTA}_S(R_t) = (1 - \delta) \alpha X_t + \delta V^{FTA}_S(R_t) + \alpha \Delta_R = \alpha X_t + \alpha \frac{\Delta_R}{1 - \delta}.
$$
Thus, at the disagreement stage at time $t$, S’s payoff is, when S conserves:

$$ (1 - \delta) q X_t + \delta V_{S}^{FTA}(R_t) = q X_t + \delta \alpha \frac{\Delta R}{1 - \delta}. $$

If, instead, S depletes, S obtains:

$$ q X_t + (a + b - c) R_t + \delta \alpha [c + (\pi - a) R_0]. $$

By comparison, S is better off conserving if:

$$ \delta \alpha \frac{\Delta R}{1 - \delta} \geq (a + b - c) R_t + \delta \alpha [c + (\pi - a) R_0] \Leftrightarrow $$

$$ \eta \geq \phi, \text{ where } \phi = \frac{a + b - c - \delta \alpha (a + b - c - d)}{\delta \alpha \left[ (c + d - a - b) + \left( \frac{b - b}{1 - \delta} \right) \right]} \Leftrightarrow (6). $$

S’s best response is a decreasing (step-)function of $\eta$; there is a unique fixed point.

**Fixed point.** If $\phi > 1$, it’s never a best response for S to conserve because $\eta > \phi$ contradicts $\eta \in [0, 1].$

In equilibrium, then, $\eta = 1$. If $\phi \leq 0$, it is always a best response to conserve because $\eta \geq \phi$ always holds; thus, $\eta = 0$. If $\phi \in (0, 1)$, the fixed point is $\eta = \phi$.

**Proposition 5.**

When S is willing to conserve after a disagreement, S’s payoff is given by (14). S’s payoff is also given by (7). Thus:

$$ \tau + \pi R_0 + (b - c) R_t = q X_t + \alpha \frac{\Delta R}{1 - \delta} \Leftrightarrow $$

$$ \tau = (c - a - b) R_t - (\pi - a) R_0 $$

$$ + \alpha \left[ c + (\pi - a) R_0 + (a + b - c - d) R_t + \left[ c + d - a - b + \frac{b - b}{1 - \delta} \right] \eta R_t \right] $$

$$ = \alpha e - (1 - \alpha) (\pi - a) R_0 - R_t \left[ \alpha d - (1 - \alpha) (c - a - b) - \alpha \left[ c + d - a - b + \frac{b - b}{1 - \delta} \right] \right]. $$

1) If $\phi \leq 0$, $\eta = 0$, so:

$$ \tau = \alpha e - (1 - \alpha) (\pi - a) R_0 - R_t \left[ \alpha d - (1 - \alpha) (c - a - b) \right]. $$

2) If $\phi = \frac{d - d}{c + d - a - b + \frac{b - b}{1 - \delta}} \in (0, 1)$, $\eta = \phi$, and:

$$ \tau = \alpha e - (1 - \alpha) (\pi - a) R_0 - R_t \left[ \alpha d - (1 - \alpha) (c - a - b) - \alpha \left[ d - \bar{d} \right] \right] $$

$$ = \alpha e - (1 - \alpha) (\pi - a) R_0 - R_t \left[ \frac{1 - \delta}{\delta} (c - b - a) + (\bar{b} - \bar{b}) \right]. $$

3) If $\phi > 1$, S’s best response is $x^D = R_t$, so:

$$ V_{S}^{FTA}(R_t) = q R_0 + (b - c) R_t + \alpha \left[ c + (\pi - a) R_0 + b - b \right]. $$

This payoff equals (7) if:

$$ \tau + \pi R_0 + (b - c) R_t = q R_0 + (b - c) R_t + \alpha \left[ c + (\pi - a) R_0 + b - b \right] \Leftrightarrow $$

$$ \tau = \alpha e - (1 - \alpha) (\pi - a) R_0 - (1 - \alpha) (b - b) R_t. $$
Lemma 1.

If N and S have signed the agreement, and S conserves at \( R_t \), S obtains:

\[ V_{SA}^{CTA}(R_t) = \pi X_t + \tau(R_t; R_T) . \]

Alternatively, if S extracts \( x_t \) to enjoy \( V_{SA}^{CTA}(R_t - x_t) \), in addition to \((b - c) x_t\), where \( \tau(R_t - x_t; R_T) \) depends on \( R_t - x_t \), S obtains:

\[ V_{SA}^{CTA}(R_t - x_t) + (b - c) x_t , \]

which is lower if

\[ V_{SA}^{CTA}(R_t) - V_{SA}^{CTA}(R_t - x_t) \geq (b - c) x_t , \]

which always holds iff:

\[ \frac{\partial V_{SA}^{CTA}(R_t)}{\partial R_t} \geq b - c , \tag{15} \]

and, note that with \( \frac{\partial V_{SA}^{CTA}(R_t)}{\partial R_t} = \frac{\partial \tau(R_t; R_T)}{\partial R_t} - \pi \), we obtain (8). Since \( \tau(0; R_T) \geq -(\pi - \frac{\partial}{\partial A}) R_0 \), to respect that S benefits from trade, integrating over (8) gives:

\[ \tau(R_t; R_T) \geq (\pi + b - c) R_t - (\pi - \frac{\partial}{\partial A}) R_0 . \tag{16} \]

Proposition 6.

Bargaining surplus. Again, let \( x_t^D = \eta R_t \), with \( \eta \in [0,1] \), measure S’s extraction after disagreement. (In principle, \( \eta \) can be a function of \( R_t \).) The proof proceeds by deriving the fixed point where S’s best response, given \( \eta \), coincides with \( \eta \).

If the parties have disagreed at \( T \), but expect to agree at \( T + 1 \), the sum of disagreement payoffs is:

\[ V^{DIS}(R_T) = (1 - \delta) (X_T + \eta R_T) a - (c + d - b) \eta R_T + \delta V_{SA}^{CTA}((1 - \eta) R_T) = (1 - \delta) (X_T + \eta R_T) a - (c + d - b) \eta R_T + \delta [e + \pi (X_T + \eta R_T)] . \]

The total gains from agreeing at \( T \), \( e + \pi X_T \), minus the above disagreement payoff, \( V^{DIS}(R_T) \), is:

\[ \Delta R = (1 - \delta) \left[ e + (\pi - \frac{\partial}{\partial A}) X_T \right] - (1 - \delta) \eta R_T a + (c + d - b) \eta R_T - \delta \pi R_T \\
= (1 - \delta) \left[ e + (\pi - \frac{\partial}{\partial A}) X_T \right] + (c + d - a - b - \delta (\pi - \frac{\partial}{\partial A})) \eta R_T . \tag{17} \]

S’s best response. Consider, first, the case in which \( x_t^D = 0 \) is among S’s best responses. Even after disagreement at \( T \), N and S will agree at \( T + 1 \), so S expects S’s default payoff plus fraction \( \alpha \) of (17):

\[ V_{SA}^{CTA}(R_T) = (1 - \delta) a X_T + \delta V_{SA}^{CTA}(R_T) + \alpha \Delta R = a X_T + \alpha \frac{\Delta R}{1 - \delta} . \tag{18} \]

Thus, after disagreeing at time \( T \), S’s payoff is, when S conserves:

\[ (1 - \delta) a X_T + \delta V_{SA}^{CTA}(R_T) = a X_T + \delta \alpha \frac{\Delta R}{1 - \delta} . \]

If, instead, S depletes, S obtains:

\[ a X_T + (a + b - c) R_T + \delta \alpha [e + (\pi - \frac{\partial}{\partial A}) R_0] . \]

S is better off conserving if:

\[ \delta \alpha \frac{\Delta R}{1 - \delta} \geq (a + b - c) R_T + \delta \alpha [e + (\pi - \frac{\partial}{\partial A}) R_0] \Leftrightarrow \\
- \delta \alpha (\pi - \frac{\partial}{\partial A}) R_T + \delta \alpha \left[ \frac{c + d - a - b - \delta (\pi - \frac{\partial}{\partial A})}{1 - \delta} \right] \eta R_T \geq (a + b - c) R_T \Leftrightarrow \\
\eta \geq \varphi , \text{ where } \varphi = \frac{1 - \delta}{\delta \alpha} \frac{a + b - c + \delta (\pi - \frac{\partial}{\partial A})}{c + d - a - b - \delta (\pi - \frac{\partial}{\partial A})} . \]

S’s best response is a decreasing (step-)function of \( \eta \); there is a unique fixed point.
Fixed point. If \( \varphi > 1 \), it is never a best response for \( S \) to conserve because \( \eta > \varphi \) contradicts \( \eta \in [0,1] \). In equilibrium, then, \( \eta = 1 \). If \( \varphi \leq 0 \), it is always a best response to conserve because \( \eta \geq \varphi \) always holds; thus, \( \eta = 0 \). If \( \varphi \in (0,1] \), the fixed point is \( \eta = \varphi \).

Tariffs. When \( S \) is willing to conserve after a disagreement, \( S \)'s payoff is given by (18). Because the CTA motivates conservation, this payoff is \( \pi X_T + \tau \). Thus:

\[
\pi X_T + \tau = a X_T + \alpha \frac{\Delta R}{1-\delta} \iff \tau = \alpha c - (1-\alpha)(\pi - a) X_T + \alpha \left( \frac{c + d - a - b - \delta (\pi - a)}{1-\delta} \right) \eta R_T.
\]

1) If \( \varphi \leq 0 \), \( \eta = 0 \), so:

\[
\tau = \alpha c - (1-\alpha)(\pi - a) X_T.
\]

2) If \( \varphi = \frac{1-\delta}{\delta a} \frac{a+b-c+\delta a(\pi-a)}{c+d-a-\delta(\pi-a)} \in (0,1] \), \( \eta = \varphi \), and then:

\[
\tau = \alpha c - (1-\alpha)(\pi - a) R_T + \frac{a+b-c}{\delta} + \pi - a R_T.
\]

3) If \( \varphi > 1 \), \( S \)'s best response after disagreement is \( x^D = R_t \), so:

\[
V^{\text{CTA}}_S(R_t) = a R_0 + (b-c) R_T + \alpha [e + (\pi-a) X_T + (c+d-a-b) R_T].
\]

This payoff equals \( \pi X_T + \tau \) if:

\[
\tau = a R_0 - \pi X_T + (b-c) R_T + \alpha [e + (\pi-a) X_T + (c+d-a-b) R_T] \\
= \alpha c - (1-\alpha)(\pi - a) R_0 + [ad + (1-\alpha)(\pi+b-c)] R_T.
\]

It is easy to check that, in all three cases, \( \tau \) satisfies (16) when \( R_t = R_T \). This implies that with the above \( \tau \), it is possible to find a function \( \tau(R_t; R_T) \) satisfying (8) for every \( R_t \in [0, R_T] \).

**Proposition 7.**

(i) If \( \alpha = 0 \), \( N \) minimizes \( V^{\text{CTA}}_S(R_T) \) s.t. the conditions that \( S \) conserves: i.e., (15), and that \( S \) accepts to trade, requiring \( V^{\text{CTA}}_S(0) \geq a R_0 \). When both bind:

\[
V^{\text{CTA}}_S(R_T) = a R_0 + (b-c) R_T.
\]

With the largest tariffs on \( e \) and \( x_t \), and the lowest on \( S \)'s beef, \( V^{\text{CTA}}_S(R_T) \leq e + \pi (R_0 - R_T) \), so:

\[
a R_0 + (b-c) R_T \leq e + \pi (R_0 - R_T) \iff (10).
\]

(ii) This part takes advantage of the fact that if the tariffs can depend on stocks, they can depend on stock differences, and \( x_t \). If \( x(R_t) \) is implemented by the agreement, \( S \)'s continuation value is:

\[
V^{\text{CTA}}_S(R_t) = x(R_t) \left( \bar{b} (1-\tau_b) - c \right) + (1-\delta) (1-\tau_S) (X_t + x(R_t)) \pi + (1-\delta) \tau_N e + \delta V^{\text{CTA}}_S(R_t - x(R_t)),
\]

where \( \tau_b \) is the tariff on timber. When \( \alpha = 0 \), \( N \) ensures that any deviation leaves \( S \) with its autarky payoff, \( a R_0 + (b-c) R_t \). \( N \)'s problem is to minimize \( x(R_t) \), s.t. \( V^{\text{CTA}}_S(R_t) \geq a R_0 + (b-c) R_t \). Thus, \( \tau_b = \tau_S = 0 \), \( \tau_N = 1 \), and \( V^{\text{CTA}}_S(R) = V^{\text{AUT}}_S(R) \). With this, \( x(R_t) \) is given by:

\[
a R_0 + (b-c) R_t = x(R_t) \left( \bar{b} - c \right) + (1-\delta) (X_t + x(R_t)) \pi + (1-\delta) e + \delta (a R_0 + (b-c) (R_t - x(R_t))) \iff (11).
\]

18
Online Appendix

Online Appendix A. Comparison to Commitment

Above, the parties are not endowed with the ability to commit: The CTA tariffs have been required to be renegotiation proof. Nevertheless, Proposition 6 shows that the CTA can implement the first-best outcome. When subsidies cannot be used, Proposition 7 describes the amount that the CTA will conserve, in equilibrium. In this situation, one might wonder if there can be other designs that can be even better and that can motivate more conservation.

The answer to this question is no. Even if the parties could commit to policies that were conditioned on the resource stock and extraction levels, they would not be able to obtain higher payoffs than from the CTA, as long as export subsidies cannot be used. To prove this claim, it is sufficient to maximize the amount of conservation subject to the harshest punishment on $S$ if $S$ deviates from the plan. The harshest punishment is autarky. The autarky payoff is also what $S$ obtains if $S$ decides to fully deplete under the CTA, and full depletion is indeed a best response for $S$ as long as marginal depletion (i.e., $x_t$ marginally larger than (11)) is a best response.\footnote{The last statement follows because $S$’s payoff is linear in the size of the stock.}

**Proposition 8.** The CTAs described by Propositions 6 and 7 implement the same outcome, and secure the same payoffs, as $N$ and $S$ would have achieved if they could commit to future policies as a function of the history.

This result is important because it suggests that the CTA is not simply a design that improves marginally on the FTA, and from which $N$ and $S$ might be able to make further improvements. Instead, the CTA here implements the best $N$ and $S$ can hope for, even if they could have committed, although the CTA does not require them to be endowed with an ability to commit.

Online Appendix B. Binding vs. Non-binding Agreements – and Implementation

So far, the CTA has been praised as renegotiation proof because it distributes all gains from trade and, therefore, no other agreement is weakly better for both parties and strictly better for one. Renegotiation proofness is a natural requirement for a treaty that is binding, that is, if the agreement binds each party unless both countries agree to renegotiate the terms. If the agreement is non-binding, however, an individual country is free to tear it apart. If a country does so, the parties will find it in their interests to agree on another Pareto optimal allocation where $S$ captures the fraction $\alpha$ of the time $t$ surplus relative to autarky. I will say that an agreement is "renego proof" if at no $t \geq T$ or $R_t \in [0, R_T]$, no party can strictly benefit from leaving the agreement (e.g., in order to negotiate a new one).\footnote{In principle, it is not clear whether this threat of sticking with autarky should be taken seriously by the opponent. After all, the country reneging harms itself unless it soon wins the war of attrition it has just initiated. The credibility of this threat will depend on the details of the bargaining structure. This ambiguity has motivated a variety of definitions of renegotiation proofness that I do not intend to survey here. The above notion of renegotiation proofness is referred to as "the standard one" by Abreu et al. (1993) and Bergin and MacLeod (1993), and these authors propose concepts that are related to renego proofness. Mailath and Sammelson (2006) review the early literature on this topic.}

The CTA, described above, can be renego proof as well as renegotiation proof. As mentioned after Proposition 6, $\tau(R_T; R_T)$ increases in $R_T$. When the CTA leads to conservation, then $N$ will pay $S$, in equilibrium, for the conservation benefits that $N$ enjoys from the CTA. If, instead, $S$ has already depleted the resource, then $S$ will obtain less favorable terms of trade because $N$ benefits less from the CTA. This fact is often sufficient to motivate $S$ to conserve the resource.

**Proposition 9.** Suppose $N$ and $S$ are free to renego on the CTA at any point in time.

(i) Suppose export subsidies are available. If $c \notin (0, \alpha (\pi - \bar{a}))$, the equilibrium CTA is $\tau(R_t; R_T) = \tau(R_t; R_t)$, where $\tau(R_t; R_t)$ is given by Proposition 6 if just $R_T$ is replaced by the current $R_t \leq R_T$. If $c \in (0, \alpha (\pi - \bar{a}))$, the equilibrium CTA is, instead:

\[
\tau(R_t; R_t) = \alpha c - (1 - \alpha) (\pi - \bar{a}) R_0 + (\pi + b - c) R_t. \tag{19}
\]

In either case, the CTA implements the first best.
(ii) If export subsidies cannot be used, and $\alpha = 1$, the equilibrium CTA is given by Proposition 7.

**Proof of Proposition 9.**

(i) If either party can walk away from the CTA and negotiate a new CTA, under the threat of autarky, then the equilibrium will be characterized by $\tau(R_t; R_T) = \tau(R_t; R_t)$, where $\tau(R_t; R_t)$ is given by Proposition 6 (where $R_T$ is replaced by the current $R_t$), if the CTA leads to conservation. This $\tau(R_t; R_t)$ increases in $R_t$, and it might satisfy Lemma 1. If we compare the derivatives $\partial \tau(R_t; R_T)/\partial R_T$ for the three cases in Proposition 6 with the requirement (8), it is easy to verify that (8) is satisfied whenever $c-a-b \notin (0, \alpha(\bar{a} - \bar{q}))$. In this case, therefore, the CTA $\tau(R_t; R_T) = \tau(R_t; R_t)$, where $\tau(R_t; R_t)$ is given by Proposition 6, is both renegotiation proof and renege proof, and it is the equilibrium treaty when the parties negotiate. No party will ever want to renege on this CTA, not even off the equilibrium path. If $c-a-b \in (0, \alpha(\bar{a} - \bar{q}))$, then $\tau(R_t; R_T)$, in Proposition 6, does not increase sufficiently fast in $R_T$ to motivate conservation. When S can renege on the CTA, the CTA will be renege proof and it will motivate S to conserve only if $\tau$ satisfies (8) for every $R_t \in [0, R_T]$. By integrating (8) from $R_t = 0$ to $R_t = R_T$, we can see that N must agree on the following $\tau$:

$$\tau(R_t; R_t) = \tau(0; 0) + (\bar{a} + b - c) R_t.$$  

Further, for $\tau(0; 0)$ to be renege proof, it must be given by Proposition 6 when $R_T = 0$. When we combine the two terms, we get (19). This CTA is renege proof, it implements the first best, and it is larger than the one in Proposition 6 if and only if $c-a-b \notin (0, \alpha(\bar{a} - \bar{q}))$.

(ii) The CTA described by Proposition 7 is renege proof by construction because N has all bargaining power and must respect S’s participation constraint at every $R_t \in [0, R_T]$. QED

When $\alpha \to 0$, it is always true that $c-a-b \notin (0, \alpha(\bar{a} - \bar{q}))$, and, thus, that the CTA characterized by Proposition 6 is renege proof. When N has all the bargaining power, it is intuitive that S cannot benefit from renegotiating. The CTA described by Proposition 7, where $\alpha = 0$ was assumed, is thus always renege proof.

**Implementation.** When the CTA is renege proof, it is straightforward to implement. It is sufficient to let parties negotiate $\tau(R_t; R_T)$ in period $T$, and allow either party renegotiate the agreement in any subsequent period, and at any point in time during that period. Therefore, it is sufficient that $\tau(R_t; R_T)$ holds through period $T$, and that it is sensitive to the stock that is relevant at the consumption stage at time $T$, that is, $R_{T+1} = R_T + x_T$. The $\tau$ for subsequent periods can be negotiated later.

**Online Appendix C. Remarks on Exhaustibility and Irreversability.**

The negative results from Proposition 2 and 3 follow because the resource is exhaustible and depletion irreversible. If, instead, $R_t$ returned to $R_0$ after every period, or if the stock was not relevant, then N and S would lose from trade if $\bar{a} + \bar{b} < c + d$, and S would not be able to extract to obtain an FTA.

The CTA, in contrast, can secure conservation because the resource is exhaustible. If, instead, $R_t$ returned to $R_0$ in every period, or if $R_t$ were not relevant, then it would not be credible that $\tau$ would decrease if S extracted. If the anticipation of such a decrease could motivate S to conserve, then N would prefer to "restart the clock" after S had extracted. For this reason, the CTA would not be renegotiation proof if the resource were renewable.

Consequently, while the exhaustibility feature intensifies the conflict between trade and conservation under the FTA, it is this feature that makes the CTA effective and credible in motivating conservation.

**References for Online Appendix**

