TRADE AND TREES

Bård Harstad

Stanford University

October 27, 2023

Abstract

International trade and natural resource exploitation interact in multiple ways. This paper first presents a dynamic game in which the South (S) exploits (e.g., deforests) in order to export (e.g., lumber or agricultural products). Because of negative externalities, the North might lose from trade, unless the resource has already been depleted. Anticipating this, S exploits more.

All negative results are reversed if renegotiation-proof tariffs can be contingent on the size of the remaining resource stock. Larger gains from trade, and more attractive terms of trade, can be used to slow exploitation. Combined with export subsidies, the outcome is first best.

I. INTRODUCTION

Trade liberalization can lead to deforestation in the tropics.[1] In 2019, Brazil led the Mercosur trade bloc to conclude its largest trade agreement ever with the European Union. At the same time, deforestation rates increased and the forest fires in the Amazon gained international media attention. The following year, France and the Netherlands proposed a "staged implementation of tariff reduction linked to the effective implementation" of sustainable provisions, including "the possibility of withdrawal of those specific tariff lines in the event of a breach of those provisions."[2]

This paper first presents a simple model that is intentionally designed to capture the negative relationships between free trade and environmental conservation. The framework is used to investigate when contingent tariffs can reverse the negative results.

The model includes the South (S), endowed with a depletable resource, and the North (N), preferring conservation. The resource can be oil, coal, or – as in my leading example – 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.

All results are reversed with 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. The difference motivates S to reduce the deforestation rate, even when the tariff functions are renegotiation proof and thus credible. If the tariffs can be accompanied by N's export subsidies, the first best can be implemented.

The CTA is not only efficient, but it is also the equilibrium outcome because it maximizes the social surplus and this surplus will be shared by N and S.

The Online Appendix shows that the equilibrium CTA implements the same outcome that N and S would have obtained if they had 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

---

[1] See Faria et al. (2016), Pendrill et al. (2019), and Abman and Lundberg (2020). Farrokhi et al. (2023) show, however, that forest cover might increase in the North.

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

further improvements: It implements the first best if export subsidies are available, and the second best 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. Section V explains that CTAs can be highly effective, according to calibrated models (Harstad, 2023) and that the design sheds light on the 2020 policy proposal by France and the Netherlands.

Hartman (1976) showed that if the standing forest has some value, it is optimal to log later. Recent research on tropical deforestation emphasize the value of conserving forever (Franklin adn Pindyck, 2018). Unfortunately, the monetary compensation program "has faced a number of challenges," according to the UN (IPBES, 2019:54). The challenges include liquidity constraints (Jayachandran, 2013), embezzlement (Caselli and Michaels, 2013), corruption (Brollo et al., 2013), and that future payments may not be credible (Harstad, 2016). Consequently, it is important to investigate how conservation can be motivated without the use of explicit payments.

When trade causes environmental problems, other scholars have analyzed trade sanctions (Barrett, 1997), border tax adjustments (Ludema and Wooton, 1994), and climate clubs (Nordhaus, 2015). Farrokhi and Lashkaripour (2021) compare the effects. 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 (Ederington, 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 approach 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).[3] 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.

II. The Model

The Players. Let the South, S, be endowed with a resource stock that can be depleted over time. At the beginning of time $t \in \{1, 2, ...\}$, the stock is R_t, and the part that has been exploited is X_t. When S exploits $x_t \in [0, R_t]$,

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

---

[3] 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 be any depletable resource, but I'll 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.

The North, N, is S's potential trading partner. The gains from trade are described below.

Timing. In each period, t, the players 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.

Payoffs. In autarky, $\underline{a} \geq 0$ represents S's (present-discounted) value of converted land, if $(1 - \delta) \underline{a}$ measures S's *per-period* utility of the beef produced per unit of land and $\delta \in (0, 1)$ is the discount factor. In addition, $\underline{b} \geq 0$ 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:

$$V_S^{AUT}(R_t) \equiv \max_{x_t^{AUT} \in [0, R_t]} u_S^{AUT}(R_t, x_t) + \delta V_S^{AUT}(R_{t+1}), \text{ where}$$

$$u_S^{AUT}(R_t, x_t) \equiv (1 - \delta)(X_t + x_t)\underline{a} + \underline{b}x_t - cx_t$$

is S's flow payoff. The linearity in x_t implies that the autarky choices are simple:

$$x_t^{AUT}(R_t) = \left\{ \begin{array}{l} 0 \text{ if } \underline{a} + \underline{b} \leq c \\ x_t \in [0, R_t] \text{ if } \underline{a} + \underline{b} = c \\ R_t \text{ if } \underline{a} + \underline{b} > c \end{array} \right\}. \tag{1}$$

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 $(1 - \delta)\bar{a}$. That is, $\bar{a}$ 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 $\bar{b}$.

It is without loss of generality to let $\bar{a} \geq \underline{a}$ and $\bar{b} \geq \underline{b}$.[4] I assume that the seller sets the price so, with a free trade agreement (FTA), S receives $(1 - \delta)\bar{a}$ for each exported unit of beef, $\bar{b}$ for each unit of timber and, hence, flow payoff

$$u_S^{FTA}(R_t, x_t) = (1 - \delta)(X_t + x_t)\bar{a} + \bar{b}x_t - cx_t.$$

For the set of goods that N can export, S's willingness to pay ($\bar{e}$) may be larger than N's willingness to pay ($\underline{e}$). Because the seller captures the gains from trade, N's flow payoff is $u_N^{AUT}(R_t, x_t) = (1 - \delta)\underline{e} - dx_t$ in autarky and $u_N^{FTA}(R_t, x_t) = (1 - \delta)\bar{e} - dx_t$ with free trade. Hence, N earns $e \equiv \bar{e} - \underline{e} \geq 0$ from exporting when trade is liberalized.

---

[4] If, for example, $\bar{a} < \underline{a}$, S's beef will not be exported and the realized gains from trade will be zero, i.e., exactly the same as when $\bar{a} = \underline{a}$.

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.

First Best. Utilities are transferable, so the gains from trade at time t are:

$$(1 - \delta) e + (1 - \delta) (\overline{a} - \underline{a}) (X_t + x_t) + (\overline{b} - \underline{b}) x_t > 0.$$

For there to be an environmental problem, assume:

$$\overline{a} + \overline{b} < c + d. \tag{2}$$

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

Bargaining Solution. Assume N and S sign an agreement if and only if it increases total payoffs, and S's share of the bargaining surplus is $\alpha \in [0, 1]$.[5] Thus, side transfers can be used if the countries agree on an FTA. Section III.E explains that negotiating tariffs can substitute for explicit 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). Section IV derives an efficient outcome when CTAs are possible, *even if* attention is restricted to Markov-perfect equilibria (MPEs) in pure strategies.

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 $x_t^{AUT} > 0$ only if $\underline{a} + \underline{b} > c$. With trade, S exploits also if $\overline{a} + \overline{b} > c$. It follows that S exploits more with trade than in autarky.

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

$$
\begin{aligned}
x_t^{FTA} &= R_t \geq x_t^{AUT} = 0, \text{ if} \\
c &\in [\underline{a} + \underline{b}, \overline{a} + \overline{b}).
\end{aligned} \tag{3}
$$

If $\overline{a} + \overline{b} \leq c$, then $x_t^{AUT} = x_t^{FTA} = 0$. If $\underline{a} + \underline{b} > c$, then $x_t^{AUT} = x_t^{FTA} = R_t$.

All proofs are in the Appendix.

---

[5] These assumptions hold if the parties use the asymmetric Nash Bargaining Solution, or if N and S make a take-it-or-leave-it offer with probabilities $1 - \alpha$ and α, respectively.

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.

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:*

$$R_t \leq R^* \equiv \frac{e + (\bar{a} - \underline{a}) R_0}{c + d - \underline{a} - \bar{b}} \tag{4}$$

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. Exploitation to Trade

Because the parties cannot commit, S anticipates that when $R_t \leq R^*$, S will be able to trade. Even when $R_t > R^*$ and $\underline{a} + \underline{b} < c$, so that S finds it costly to exploit in autarky, this cost is worth paying if the required exploitation $(R_t - R^*)$ is small or 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:*

$$R_t < \widehat{R} \equiv \delta \alpha \frac{e + (\bar{a} - \underline{a}) R_0}{c - \underline{a} - \underline{b}} \ \text{ or } \ \underline{a} + \underline{b} > c. \tag{5}$$

By comparison,

$$\widehat{R} > R^* \Leftrightarrow d > \widehat{d} \equiv \left(\frac{1}{\delta \alpha} - 1 \right) (c - \underline{a} - \underline{b}) + \bar{b} - \underline{b}.$$

When $R_t \in \left(R^*, \widehat{R} \right)$, 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 (α) 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 $R_t \in \left(R^*, \widehat{R} \right)$. There are multiple SPEs, but all lead to trade and exploitation.[6]

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 the unique MPE in linear strategies: The FTA is signed, and S exploits, iff the gains from trade are large or R_t is small:*

$$if \ d \le \widehat{d}, \ the \ condition \ is \ \frac{e + (\overline{a} - \underline{a}) R_0}{R_t} > c + d - \underline{a} - \overline{b}, \ and \ x^D = 0,$$

$$if \ d > \widehat{d}, \ the \ condition \ is \ \frac{e + (\overline{a} - \underline{a}) R_0}{R_t} > \frac{c - \underline{a} - b}{\delta \alpha}, \ and \ x^D = R_t \cdot \min \{\phi, 1\},$$

$$where \ \phi \equiv \frac{d - \widehat{d}}{c + d - \underline{a} - \overline{b} + \frac{\overline{b} - \underline{b}}{1 - \delta}}. \tag{6}$$

In region A in Fig. 1, $d < \widehat{d}$, and S is not willing to exploit in order to obtain an FTA when (4) fails. In this situation, N and S trade *iff* $R_t \le R^*$, 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 $\phi > 0$. This fraction may be less than 1 because if x^D were very large, the bargaining surplus from signing the FTA would be large, and – anticipating the fraction α of it – S would be willing to conserve; a contradiction. If, in equilibrium, $\phi \in (0, 1)$, S is indifferent, as in a mixed-strategy equilibrium.[7]

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 α 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

---

[6] For example, there are SPEs in which N and S agree to trade at some future time, or in every Δ_t period, where Δ_t is so small that S prefers to conserve when there are fewer than Δ_t periods left.

[7] In the mixed-strategy equilibria of this game, S's expected default extraction level is ϕR_t, given by (6).
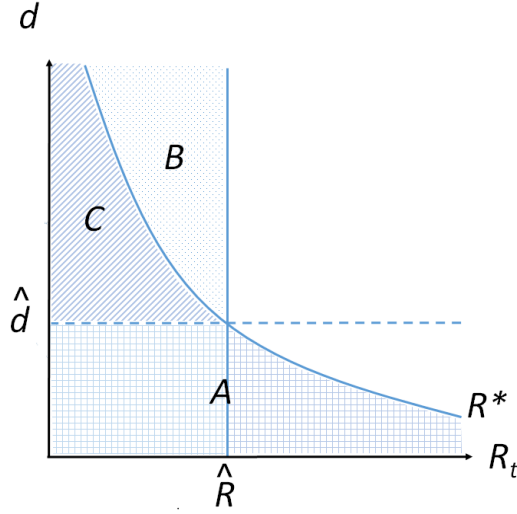
Figure 1: *An FTA is socially optimal if $R_t < R^*$, but S is willing to exploit to obtain one if $R_t < \widehat{R}$. In region B, N and S trade even if the total surplus is larger in autarky.*

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

III.E.   Tariffs as Transfers

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 $(\bar{a} - \underline{a}) R_0/R_t > c - \underline{a} - \underline{b}$.

Here, transfers can take the form of tariff adjustments. If $\tau_N \leq 1$ measures N's ad valorem tariff on imported beef, then consumers in N are willing to pay only $(1 - \tau_N)\bar{a}$ for S's beef. S will find it necessary (and optimal) to reduce the price and, therefore, S loses and N gains.

A tariff (τ_S) in S on S's import is similarly improving S's, but worsening N's, terms of trade.

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

$$
\begin{aligned}
V_S^{FTA}(R_t, \bar{\tau}) &= \tau_S \bar{e} + (1 - \tau_N)\bar{a}R_0 + (\bar{b} - c) R_t = V_S^{FTA}(R_t, 0) + \bar{\tau}, \text{ and} \\
V_N^{FTA}(R_t, \bar{\tau}) &= (1 - \tau_S)\bar{e} + \tau_N \bar{a}R_0 - dR_t = V_N^{FTA}(R_t, 0) - \bar{\tau}, \text{ where} \\
\bar{\tau} &\equiv \tau_S \bar{e} - \tau_N \bar{a}R_0.
\end{aligned}
\tag{7}
$$

Note that $\bar{\tau}$ essentially represents a transfer from N to S.[8]

If there are no export subsidies, S is willing to export only if $(1 - \tau_N)\bar{a} \geq \underline{a}$, and N if $(1 - \tau_S)\bar{e} \geq \underline{e}$. These constraints can be ignored if the tariffs can be accompanied by export subsidies (Grossman and

[8] To facilitate this transfer, there is no need to introduce tariffs on timber (x_t) as well.

Helpman, 1995): If s_N is an ad valorem export subsidy in N, N's producers are willing to export as long as $(1 + s_N)(1 - \tau_S)\bar{e} \geq \underline{e}$. Alternatively, the transfers can be arbitrarily large if, instead of export subsidies, we permit import subsidies.

The equilibrium transfer from N to S will reflect the bargaining strength (α), the gains from trade, and the payoffs in the outside option (i.e., in autarky).

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.

PROPOSITION 5. *If N and S sign an FTA at time T, then τ_S is smaller or τ_N is larger if R_T is large:*

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

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, τ_S must decrease in R_T, or τ_N 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 (τ_S, τ_N), every such allocation is renegotiation proof in the following sense: any change in $\bar{\tau}$ 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 (τ_S, τ_N), 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 $\tau_N(R_t; R_T)$ and $\tau_S(R_t; R_T)$, 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. When $R_t \leq R^*$, it is always (ex post) optimal for N and S to trade, according to Proposition 2. 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

A challenge is that, even if there is no exploitation after the agreement is signed, the sum of the gains from trade, $e + (\bar{a} - \underline{a}) X_t$, 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 $\bar{\tau}$ increases sufficiently fast in R_t, then S has an incentive to conserve rather than to exploit.

LEMMA 1. *S conserves by selecting* $x_t = 0 \forall t \geq T$ *iff:*

$$\partial \frac{\bar{\tau}(R_t; R_T)}{\partial R_t} \geq \bar{a} + \underline{b} - c \ \forall R_t \leq R_T, \ where \tag{8}$$
$$\bar{\tau}(R_t; R_T) \equiv \tau_S(R_t; R_T) \bar{e} - \tau_N(R_t; R_T) \bar{a} X_t.$$

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 ($\bar{\tau}$ 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 the unique MPE in linear strategies:*
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:

$$\bar{\tau}(R_T; R_T) = \left\{ \begin{array}{ll} \alpha \bar{e} - (1-\alpha)(\bar{a}-\underline{a}) R_0 + (1-\alpha)(\bar{a}-\underline{a}) R_T & and \ x^D = 0 \quad if \ \varphi < 0 \\ \alpha \bar{e} - (1-\alpha)(\bar{a}-\underline{a}) R_0 + \frac{a+b-c}{\delta} R_T + (\bar{a}-\underline{a}) R_T & and \ x^D = \varphi R_T \ if \ \varphi \in (0,1) \\ \alpha \bar{e} - (1-\alpha)(\bar{a}-\underline{a}) R_0 + [(1-\alpha)(\bar{a}+\underline{b}-c)+\alpha d] R_T & and \ x^D = R_T \quad if \ \varphi > 1 \end{array} \right\}$$

$$where \ \varphi \equiv \frac{\delta \alpha (\bar{a}-\underline{a}) + a + \underline{b} - c}{\delta \alpha (\bar{a}-\underline{a}) + \delta \alpha \frac{d+c-\bar{a}-\underline{b}}{1-\delta}}.$$

There are two reasons for why $\overline{\tau}$ increases in R_T: N's gain from conservation, and thus the CTA, is larger if R_T is large; and S's gains from trade is smaller if less land has been deforested.

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 τ_S to be so large that N must subsidize its export for the producers to be willing to sell, or that $\tau_N < 0$, so that N subsidizes import. When τ_S is accompanied by an export subsidy in country N, or N subsidizes import, then there is no limit to how large $\overline{\tau}$ can be, and there is no limit to how much one can conserve. N agrees to the large $\overline{\tau}$, in equilibrium, because it motivates conservation.

IV.C. Reducing 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 the unique MPE in linear strategies:*
N offers a CTA that S immediately accepts:
(i) The tariffs are as in Proposition 6 and $x_t = 0$ for every $t \geq T$ if:

$$\frac{e + (\overline{a} - \underline{a}) R_0}{R_T} \geq \overline{a} + \underline{b} - c \Leftrightarrow \tag{9}$$

$$R_T \leq \widetilde{R} \equiv \frac{e + (\overline{a} - \underline{a}) R_0}{\overline{a} + \underline{b} - c} \ or \ \overline{a} + \underline{b} < c. \tag{10}$$

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

$$x_t = \left(R_t - \widetilde{R} \right) \gamma, \ where \ \gamma \equiv \frac{\overline{a} + \underline{b} - c}{\overline{a} + \underline{b} - c + \frac{\overline{b} - \underline{b}}{1 - \delta}} \in (0, 1), \tag{11}$$

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

Eq. (9) can be rewritten as:

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

With this reformulation, part (i) in Proposition 7 states that if $\underline{a} + \underline{b} < c$, 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 $\underline{a} + \underline{b} > c$, 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, $\widetilde{R}$, is limited by the value of trade.
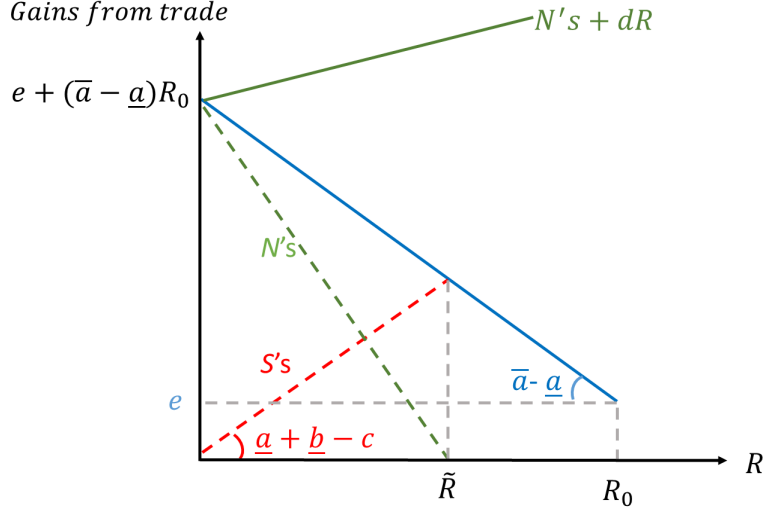
Figure 2: *When S's gains from trade is limited by the total gains from trade, which decreases in R (solid line), at most $\widetilde{R}$ can be conserved. When S's gain must increase in R to motivate conservation, N's gain must decline in R (dashed line). N's total gain is, nevertheless, increasing in R when the avoided damage is taken into account (upward-sloping solid line).*

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 $\widetilde{R}$.

Part (ii) states that if $R_T > \widetilde{R} > 0$, 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 the gains from trade are, the more it is possible to persuade S to conserve in each period.

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

REMARK ON α. If $\alpha > 0$, the equilibrium x_t, when $R_T > \widetilde{R}$, would be larger than the x_t described by (11). When $R_T > \widetilde{R}$, the level of α is thus not only affecting the distribution of surplus, but also efficiency: a larger α 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.

COROLLARIES. *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) τ_S strictly increases or τ_N strictly decreases (or both) in R_T as well as in R_t.

Compared to the FTA, the CTA implements strictly higher payoffs as long as the first best requires conservation but free trade leads to exploitation, i.e., under:

$$\overline{a} + \overline{b} - c - d < 0 < \overline{a} + \overline{b} - c. \tag{12}$$

V. Conclusion

The model above is designed to illustrate that free trade can cause environmental exploitation, exploitation can cause trade, and countries might exploit to obtain a free trade agreement. Nevertheless, the negative results are reversed and exploitation can be postponed with a contingent trade agreement.

The model is extremely stylized, but can be generalized: With nonlinear utility functions, the optimal levels of deforestation and export can be interior and tariffs distortionary; instead of autarky, noncooperative equilibrium tariffs represent a more reasonable treat point; the model should permit multiple goods and countries. These extensions are developed in a static model (Harstad, 2023), which confirms that a CTA is *always* strictly better than an FTA. (With linear utilities, the CTA is strictly better *iff* (12) holds.) When that model is calibrated, the estimates suggest that, even if the South has a growing number of trading partners, a CTA offered by the EU and the US can end or limit deforestation.

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 analyses can shed light on how one might succeed with the 2020 policy initiative from France and the Netherlands, mentioned in the Introduction. The so-called "non-paper" concludes by arguing that the "WTO should offer an enabling space to apply sustainability disciplines." If the WTO permits export subsidies for conservation purposes, then more forests can indeed be conserved, according to the analysis above.

REFERENCES

Abman, Ryan, and Lundberg, Clark (2020): "Does Free Trade Increase Deforestation? The Effects of Regional Trade Agreements," Journal of the Association of Environmental and Resource Economists 7(1): 35-72.

Brollo, Fernanda; Nannicini, Tommaso; Perotti, Roberto, and Tabellini, Guido (2013): "The Political Resource Curse," American Economic Review 103(5): 1759-96.

Caselli, Francesco and Michaels, Guy (2013): "Do Oil Windfalls Improve Living Standards? Evidence from Brazil " American Economic Journal: Applied Economics 5(1): 208-38.

Copeland, Brian R.; Shapiro, Joseph S.; and Taylor, Scott M. (2022): "Globalization and the environment," in G. Gopinath, E. Helpman, and K. Rogoff (eds.), Handbook of International Economics 5: 61-146. Elsevier.

Ederington, Josh (2002): "Trade and Domestic Policy Linkage in International Agreements," International Economic Review 43(4): 1347-68.

Faria, Weslem Rodrigues, and Almeida, Alexandre Nunes (2016): "Relationship between openness to trade and deforestation: Empirical evidence from the Brazilian Amazon," Ecological Economics 121: 85-97.

Farrokhi, Farid; Kang, Elliot; Pellegrina, Heitor S., and Sotelo, Sebastian (2023): "Deforestation: A Global and Dynamic Perspective." Mimeo, University of Michigan.

Farrokhi, Farid, and Lashkaripour, Ahmad (2021): "Can Trade Policy Mitigate Climate Change?" Mimeo, Purdue University.

Franklin Jr, Sergio L., and Pindyck, Robert S. (2018): "Tropical Forests, Tipping Points, and the Social Cost of Deforestation," Ecological Economics 153: 161-71.

Grossman, Gene M., and Helpman, Elhanan (1995): "Trade wars and trade talks," Journal of Political Economy 103(4): 675-08.

Harstad, Bård (2016): "The Market for Conservation and Other Hostages," Journal of Economic Theory 166(November): 124-51.

Harstad, Bård (2023): "Contingent Trade Agreements." Mimeo, Stanford University.

Hartman, Richard (1976): "The Harvesting Decision When a Standing Forest Has Value," Economic Inquiry 14(1): 52-58.

Hsiao, Allan (2022): "Coordination and Commitment in International Climate Action: Evidence from Palm Oil," mimeo, Princeton University.

IPBES (2019): Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. E. S. Brondizio, J. Settele, S. Díaz, and H. T. Ngo (editors). IPBES secretariat, Bonn, Germany.

Jayachandran, Seema (2013): "Liquidity Constraints and Deforestation: The Limitations of Payments for Ecosystem Services," American Economic Review, Papers and Proceedings, 103(3): 309-13.

Limão, Nuno (2005): "Trade policy, cross-border externalities and lobbies: do linked agreements enforce more cooperative outcomes?" Journal of International Economics 67(1): 175-199.

Ludema, Rodney D., and Wooton, Ian (1994): "Cross-Border Externalities and Trade Liberalization: The Strategic Control of Pollution," Canadian Journal of Economics 27(4): 950-66.

Maggi, Giovanni (2016): "Issue linkage," in K. Bagwell and R. Staiger (Eds.), Handbook of Commercial Policy 1: 513-64.

Mailath, George J., and Samuelson, Larry (2006): Repeated games and reputations: long-run relationships. Oxford University Press.

Nordhaus, William (2015): "Climate clubs: Overcoming free-riding in international climate policy," American Economic Review 105(4): 1339-70.

Pendrill, Florence; Persson, U. Martin; Godar, Javier; Kastner, Thomas; Moran, Daniel; Schmidt, Sarah, and Wood, Richard (2019): "Agricultural and forestry trade drives large share of tropical deforestation emissions," Global Environmental Change 56: 1-10.

APPENDIX: PROOFS

Proposition 1.

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

$$
\begin{aligned}
V_S^{FTA}(0) &= V_S^{AUT}(0) + \alpha\left[e + (\bar{a} - \underline{a})R_0\right] = \underline{a}R_0 + \alpha e + \alpha(\bar{a} - \underline{a})R_0, \text{ because} \\
V_S^{AUT}(0) &= \underline{a}R_0.
\end{aligned}
$$

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

$$
V_S^{FTA}(R_t) \equiv \max_{x_t^{FTA} \in [0, R_t]} (1 - \delta)(X_t + x_t)\bar{a} + \bar{b}x_t - cx_t + \delta V_S(R_{t+1}),
$$

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

$$
x_t^{FTA}(R_t) = \left\{
\begin{array}{l}
0 \text{ if } \bar{a} + \bar{b} \leq c \\
x_t \in [0, R_t] \text{ if } \bar{a} + \bar{b} = c \\
R_t \text{ if } \bar{a} + \bar{b} > c
\end{array}
\right\}.
$$

Proposition 2.

With (3), autarky leads to total payoff $V^{AUT}(R_t) \equiv V_N^{AUT}(R_t) + V_S^{AUT}(R_t) = \underline{e} + (R_0 - R_t)\underline{a}$, while the FTA leads to depletion and total payoff

$$
V^{FTA}(R_t) \equiv V_N^{FTA}(R_t) + V_S^{FTA}(R_t) = \bar{e} + \bar{a}R_0 + (\bar{b} - c - d)R_t, \tag{13}
$$

which is larger if:

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

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:

$$
(1 - \delta)R_0\underline{a} + (\underline{b} - c)R_t + \delta V_S^{FTA}(0) > \underline{a}(R_0 - R_t) \Leftrightarrow (5).
$$

Proposition 4.

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

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

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

$$
\begin{aligned}
V^{DIS}(R_t) &= (1 - \delta)\left[\underline{e} + (X_t + \eta R_t)\underline{a}\right] - (c + d - b)\eta R_t + \delta V^{FTA}((1 - \eta)R_t) \\
&= (1 - \delta)\left[\underline{e} + (R_0 - R_t)\underline{a}\right] - \left[(1 - \delta)(c + d - \underline{a} - \bar{b}) + (\bar{b} - \underline{b})\right]\eta R_t + \delta\left[\bar{e} + \bar{a}R_0 + (\bar{b} - c - d)R_t\right].
\end{aligned}
$$

The total gains from agreeing at t, (13), minus disagreement payoff $V^{DIS}(R_t)$, is:

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

$$
= (1 - \delta)\left[e + (\bar{a} - \underline{a})R_0 + (\underline{a} + \bar{b} - c - d)R_t\right] + \left[(1 - \delta)(c + d - \underline{a} - \bar{b}) + (\bar{b} - \underline{b})\right]\eta R_t. \tag{14}
$$

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 $t + 1$ and, then, S expects S's default payoff plus fraction α of (14):

$$
V_S^{FTA}(R_t) = (1 - \delta)\underline{a}X_t + \delta V_S^{FTA}(R_t) + \alpha\Delta_R = \underline{a}X_t + \alpha\frac{\Delta_R}{1 - \delta}. \tag{15}
$$

Thus, at the disagreement stage at time t, S's payoff is, when S conserves:

$$(1 - \delta)\,\underline{a}X_t + \delta V_S^{FTA}(R_t) = \underline{a}X_t + \delta\alpha\frac{\Delta_R}{1 - \delta}.$$

If, instead, S depletes, S obtains:

$$\underline{a}X_t + (\underline{a} + \underline{b} - c)\,R_t + \delta\alpha\left[e + (\bar{a} - \underline{a})\,R_0\right].$$

By comparison, S is better off conserving if:

$$\delta\alpha\frac{\Delta_R}{1 - \delta} \geq (\underline{a} + \underline{b} - c)\,R_t + \delta\alpha\left[e + (\bar{a} - \underline{a})\,R_0\right] \Leftrightarrow$$

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

S's best response is a decreasing (step-)function of η; 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 (15). S's payoff is also given by (7). Thus:

$$\bar{\tau} + \bar{a}R_0 + \left(\bar{b} - c\right)R_t = \underline{a}X_t + \alpha\frac{\Delta_R}{1 - \delta} \Leftrightarrow$$

$$\begin{aligned}
\bar{\tau} &= \left(c - \underline{a} - \bar{b}\right)R_t - (\bar{a} - \underline{a})R_0 \\
&\quad + \alpha\left[e + (\bar{a} - \underline{a})R_0 + \left(\underline{a} + \bar{b} - c - d\right)R_t + \left[c + d - \underline{a} - \bar{b} + \frac{\bar{b} - \underline{b}}{1 - \delta}\right]\eta R_t\right] \\
&= \alpha e - (1 - \alpha)(\bar{a} - \underline{a})R_0 - R_t\left[\alpha d - (1 - \alpha)\left(c - \underline{a} - \bar{b}\right) - \alpha\eta\left[c + d - \underline{a} - \bar{b} + \frac{\bar{b} - \underline{b}}{1 - \delta}\right]\right].
\end{aligned}$$

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

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

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

$$\begin{aligned}
\bar{\tau} &= \alpha e - (1 - \alpha)(\bar{a} - \underline{a})R_0 - R_t\left[\alpha d - (1 - \alpha)\left(c - \underline{a} - \bar{b}\right) - \alpha\left(d - \widehat{d}\right)\right] \\
&= \alpha e - (1 - \alpha)(\bar{a} - \underline{a})R_0 - R_t\left[\frac{1 - \delta}{\delta}\left(c - \underline{b} - \underline{a}\right) + \left(\bar{b} - \underline{b}\right)\right].
\end{aligned}$$

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

$$V_S^{FTA}(R_t) = \underline{a}R_0 + (\underline{b} - c)R_t + \alpha\left[e + (\bar{a} - \underline{a})R_0 + \bar{b} - \underline{b}\right].$$

This payoff equals (7) if:

$$\begin{aligned}
\bar{\tau} + \bar{a}R_0 + \left(\bar{b} - c\right)R_t &= \underline{a}R_0 + (\underline{b} - c)R_t + \alpha\left[e + (\bar{a} - \underline{a})R_0 + \bar{b} - \underline{b}\right] \Leftrightarrow \\
\bar{\tau} &= \alpha e - (1 - \alpha)(\bar{a} - \underline{a})R_0 - (1 - \alpha)\left(\bar{b} - \underline{b}\right)R_t.
\end{aligned}$$

16

Lemma 1.

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

$$V_S^{CTA}(R_t) = \bar{a}X_t + \bar{\tau}(R_t; R_T).$$

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

$$V_S^{CTA}(R_t - x_t) + (\underline{b} - c)x_t,$$

which is lower if

$$V_S^{CTA}(R_t) - V_S^{CTA}(R_t - x_t) \geq (\underline{b} - c)x_t,$$

which always holds *iff*:

$$\partial \frac{V_S^{CTA}(R_t)}{\partial R_t} \geq \underline{b} - c, \tag{16}$$

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

$$\bar{\tau}(R_t; R_T) \geq (\bar{a} + \underline{b} - c)R_t - (\bar{a} - \underline{a})R_0. \tag{17}$$

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, η can be a function of R_t.) The proof proceeds by deriving the fixed point where S's best response, given η, coincides with η.

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

$$\begin{aligned} V^{DIS}(R_T) &= (1 - \delta)[\underline{e} + (X_T + \eta R_T)\underline{a}] - (c + d - \underline{b})\eta R_T + \delta V^{CTA}((1 - \eta)R_T) \\ &= (1 - \delta)[\underline{e} + (X_T + \eta R_T)\underline{a}] - (c + d - \underline{b})\eta R_T + \delta[\bar{e} + \bar{a}(X_T + \eta R_T)]. \end{aligned}$$

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

$$\Delta_R = (1 - \delta)[e + (\bar{a} - \underline{a})X_T] - (1 - \delta)\eta R_T \underline{a} + (c + d - \underline{b})\eta R_T - \delta \bar{a}\eta R_T$$

$$= (1 - \delta)[e + (\bar{a} - \underline{a})X_T] + (c + d - \underline{a} - \underline{b} - \delta(\bar{a} - \underline{a}))\eta R_T. \tag{18}$$

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 α of (18):

$$V_S^{CTA}(R_T) = (1 - \delta)\underline{a}X_T + \delta V_S^{CTA}(R_T) + \alpha \Delta_R = \underline{a}X_T + \alpha \frac{\Delta_R}{1 - \delta}. \tag{19}$$

Thus, after disagreeing at time T, S's payoff is, when S conserves:

$$(1 - \delta)\underline{a}X_T + \delta V_S^{CTA}(R_T) = \underline{a}X_T + \delta \alpha \frac{\Delta_R}{1 - \delta}.$$

If, instead, S depletes, S obtains:

$$\underline{a}X_T + (\underline{a} + \underline{b} - c)R_T + \delta \alpha[e + (\bar{a} - \underline{a})R_0].$$

S is better off conserving if:

$$\delta \alpha \frac{\Delta_R}{1-\delta} \geq (\underline{a} + \underline{b} - c)R_T + \delta \alpha[e + (\bar{a} - \underline{a})R_0] \Leftrightarrow$$

$$-\delta \alpha(\bar{a} - \underline{a})R_T + \delta \alpha \frac{(c + d - \underline{a} - \underline{b} - \delta(\bar{a} - \underline{a}))}{1 - \delta}\eta R_T \geq (\underline{a} + \underline{b} - c)R_T \Leftrightarrow$$

$$\eta \geq \varphi, \text{ where } \varphi \equiv \frac{1 - \delta}{\delta \alpha} \frac{\underline{a} + \underline{b} - c + \delta \alpha(\bar{a} - \underline{a})}{c + d - \underline{a} - \underline{b} - \delta(\bar{a} - \underline{a})}.$$

S's best response is a decreasing (step-)function of η; 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 (19). Because the CTA motivates conservation, this payoff is $\bar{a}X_T + \bar{\tau}$. Thus:

$$\bar{a}X_T + \bar{\tau} = \underline{a}X_T + \alpha \frac{\Delta_R}{1 - \delta} \Leftrightarrow \bar{\tau} = \alpha e - (1 - \alpha)(\bar{a} - \underline{a})X_T + \alpha \left(\frac{c + d - \underline{a} - \underline{b} - \delta(\bar{a} - \underline{a})}{1 - \delta}\right) \eta R_T.$$

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

$$\bar{\tau} = \alpha e - (1 - \alpha)(\bar{a} - \underline{a})X_T.$$

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

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

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

$$V_S^{CTA}(R_T) = \underline{a}R_0 + (\underline{b} - c)R_T + \alpha\left[e + (\bar{a} - \underline{a})X_T + (c + d - \underline{a} - \underline{b})R_T\right].$$

This payoff equals $\bar{a}X_T + \bar{\tau}$ if:

$$
\begin{aligned}
\bar{\tau} &= \underline{a}R_0 - \bar{a}X_T + (\underline{b} - c)R_T + \alpha\left[e + (\bar{a} - \underline{a})X_T + (c + d - \underline{a} - \underline{b})R_T\right] \\
&= \alpha e - (1 - \alpha)(\bar{a} - \underline{a})R_0 + \left[\alpha d + (1 - \alpha)(\bar{a} + \underline{b} - c)\right]R_T.
\end{aligned}
$$

It is easy to check that, in all three cases, $\bar{\tau}$ satisfies (17) when $R_t = R_T$. This implies that with the above $\bar{\tau}$, it is possible to find a function $\bar{\tau}(R_t; R_T)$ satisfying (8) for every $R_t \in [0, R_T]$.

Proposition 7.

(i) If $\alpha = 0$, N minimizes $V_S^{CTA}(R_T)$ s.t. the conditions that S conserves: i.e., (16), and that S accepts to trade, requiring $V_S^{CTA}(0) \geq \underline{a}R_0$. When both bind:

$$V_S^{CTA}(R_T) = \underline{a}R_0 + (\underline{b} - c)R_T.$$

With the largest tariffs on trees and N's export, and the lowest on S's beef, $V_S^{CTA}(R_T) \leq e + \bar{a}(R_0 - R_T)$, so:

$$\underline{a}R_0 + (\underline{b} - c)R_T \leq e + \bar{a}(R_0 - R_T) \Leftrightarrow (10).$$

(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:

$$V_S^{CTA}(R_t) = x(R_t)\left(\bar{b}(1 - \tau_b) - c\right) + (1 - \delta)(1 - \tau_N)(X_t + x(R_t))\bar{a} + (1 - \delta)\tau_S\bar{e} + \delta V_S^{CTA}(R_t - x(R_t)),$$

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

$$\underline{a}R_0 + (\underline{b} - c)R_t = x(R_t)(\bar{b} - c) + (1 - \delta)(X_t + x(R_t))\bar{a} + (1 - \delta)e + \delta(\underline{a}R_0 + (\underline{b} - c)(R_t - x(R_t))) \Leftrightarrow (11).$$

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.[9]

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 α of the time t surplus relative to autarky. I will say that an agreement is "renege 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).[10]

The CTA, described above, can be renege proof as well as renegotiation proof. As mentioned after Proposition 6, $\overline{\tau}(R_T; R_T)$ 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 renege on the CTA at any point in time.*
(i) Suppose export subsidies are available. If $c \notin (0, \alpha(\overline{a} - \underline{a}))$, the equilibrium CTA is $\overline{\tau}(R_t; R_T) = \overline{\tau}(R_t; R_t)$, where $\overline{\tau}(R_t; R_t)$ is given by Proposition 6 if just R_T is replaced by the current $R_t \leq R_T$. If $c \in (0, \alpha(\overline{a} - \underline{a}))$, the equilibrium CTA is, instead:

$$\overline{\tau}(R_t; R_t) = \alpha e - (1 - \alpha)(\overline{a} - \underline{a}) R_0 + (\overline{a} + \underline{b} - c) R_t. \tag{20}$$

In either case, the CTA implements the first best.

---

[9] The last statement follows because S's payoff is linear in the size of the stock.

[10] 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 renege proofness. Mailath and Samuelson (2006) review the early literature on this topic.

(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 $\bar{\tau}(R_t; R_T) = \bar{\tau}(R_t; R_t)$, where $\bar{\tau}(R_t; R_t)$ is given by Proposition 6 (where R_T is replaced by the current $R_t \leq R_T$), if the CTA leads to conservation. This $\bar{\tau}(R_t; R_t)$ increases in R_t, and it might satisfy Lemma 1. If we compare the derivatives $\partial \bar{\tau}(R_T; R_T)/\partial R_T$ for the three cases in Proposition 6 with the requirement (8), it is easy to verify that (8) is satisfied whenever $c - \underline{a} - \underline{b} \notin (0, \alpha(\bar{a} - \underline{a}))$. In this case, therefore, the CTA $\bar{\tau}(R_t; R_T) = \bar{\tau}(R_t; R_t)$, where $\bar{\tau}(R_t; R_t)$ 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 $c - \underline{a} - \underline{b} \in (0, \alpha(\bar{a} - \underline{a}))$, then $\bar{\tau}(R_T; R_T)$, 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 $\bar{\tau}$ 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 $\bar{\tau}$:

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

Further, for $\bar{\tau}(0; 0)$ to be renege proof, it must be given by Proposition 6 when $R_T = 0$. When we combine the two terms, we get (20).This CTA is renege proof, it implements the first best, and it is larger than the one in Proposition 6 if and only if $c - \underline{a} - \underline{b} \in (0, \alpha(\bar{a} - \underline{a}))$.

(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 $c - \underline{a} - \underline{b} \notin (0, \alpha(\bar{a} - \underline{a}))$, 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 reneging. 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 $\bar{\tau}(R_t; R_T)$ in period T, and allow either party renege on the agreement in any subsequent period, and at any point in time during that period. Therefore, it is sufficient that $\bar{\tau}(R_t; R_T)$ 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 $\bar{\tau}$ 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 $\bar{\tau}$ 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

Abreu, Dilip; Pearce, David, and Stacchetti, Ennio (1993): "Renegotiation and symmetry in repeated games," Journal of Economic Theory 60(2): 217-40.

Bergin, James, and MacLeod, Bentley (1993): "Efficiency and renegotiation in repeated games," Journal of Economic Theory 61(1): 42-73.

Mailath, George J., and Samuelson, Larry (2006): Repeated games and reputations: long-run relationships. Oxford University Press.