Multilateral Tariff Cooperation under Fairness and Reciprocity*

Costas Hadjiyiannis† Doruk İriş‡ and Chrysostomos Tabakis§

September 27, 2010

Abstract

This paper explores the impact of fairness and reciprocity on multilateral tariff cooperation. Reciprocal countries reward kind behavior (positive reciprocity), but retaliate against countries behaving unkindly (negative reciprocity). We demonstrate that reciprocal countries that are moderately demanding from their trading partners regarding their commercial policy can support a greater degree of cooperation than self-interested ones. However, when only very liberal import policies are considered fair, then reciprocity could have a detrimental effect on multilateral tariff cooperation. Thus, our model provides a novel reason for the occasional failure of trade negotiations. Finally, we show that these results are robust to endogenizing fair-tariff perceptions.

Keywords: Reciprocity; Trade agreements; Trade policy; Repeated games

JEL Classification: F13; D63

---

*We would like to thank Luis Santos-Pinto, and participants at CRETE 2009, ETSG 2009, and ASSET 2009 for very helpful comments and suggestions. A previous version of the paper was written while İriş was visiting UC San Diego, and Tabakis was an International Faculty Fellow at the MIT Sloan School of Management. Tabakis gratefully acknowledges financial support from the Fundação para a Ciência e a Tecnologia in Portugal. Any errors are ours.

†Department of Economics, University of Cyprus.
‡Faculdade de Economia, Universidade Nova de Lisboa.
§Corresponding author: Faculdade de Economia, Universidade Nova de Lisboa, Campus de Campolide, Lisbon 1099-032, Portugal; Email: ctabakis@fe.unl.pt; Phone: +351 21 380 1639; Fax: +351 21 387 0933.
1 Introduction

"We wish to do it [promote commerce] by throwing open all the doors of commerce and knocking off all its shackles. But as this cannot be done for others, unless they will do it for us, and there is no probability that Europe will do this, I suppose we may obliged to adopt a system which may shackle them in our ports, as they do to us in theirs." – Thomas Jefferson, 1785

Reciprocity is an old theme in international trade negotiations. In this paper, we set out to explore the implications of reciprocal preferences for commercial policy and multilateral trade agreements. Governments with reciprocal preferences reward "kind" or "fair" actions (positive reciprocity), whereas they punish "unkind" or "unfair" behavior (negative reciprocity). This is an important question for two reasons. First, governments seem to exhibit such preferences, at least with respect to trade policy. Second, our analysis provides a novel perspective on the successes and the occasional failures of multilateral trade negotiations. We should stress here that our definition of "reciprocity" differs substantially from the standard one in the trade literature (e.g., Bagwell and Staiger, 1999a; Freund, 2003; Krishna and Mitra, 2005). In these papers, the term "reciprocity" refers broadly to mutual changes in trade policy by self-interested countries which bring about changes in each country’s import volume that are of equal value to the changes in its export volume. Instead, in our framework, reciprocity characterizes the preferences of countries.

Most governments specifically state that one of their major goals when sitting at the negotiations table is to promote fair trade through reciprocity, with "fair trade" typically entailing that (i) domestic producers (and workers) are faced with fair trade policies worldwide; and (ii) the gains from trade are fairly distributed among trading partners. For instance, President Obama’s 2009 Trade Policy Agenda Report states that "If we work together, free and fair trade...will be a powerful contributor to the national and global well

\footnote{PTJ 8: 633.}
being." Analogous goals and concerns characterize the commercial policy of the European Union (EU). On its official website on external trade, it is written that "The EU has evolved during the process of globalization by aiming for the harmonious development of world trade and fostering fairness..." At the same time, the emphasis placed on reciprocity is equally strong. For example, President Sarkozy and Chancellor Merkel in a joint letter to the President of the EU Council in 2007 argue that "Open markets can only develop their full potential if transparent rules facilitate fair competition in a spirit of reciprocity." Moreover, a 2006 Communication of the EU Commission warns that "If necessary, targeted restrictions will be maintained [on behalf of the EU] for uncooperative countries with the aim of encouraging them towards a mutual opening up of markets." Another such example of negative reciprocity is the extensive employment of anti-dumping and countervailing measures within the context of the World Trade Organization (WTO), aimed directly at punishing unfair trade practices. In brief, in the international trade arena, governments seem, to some extent, to not be simply maximizers of their own self-interested welfare, but to rather be exhibiting some preferences for fairness and reciprocity.

It could also be argued that it is reasonable to expect governments to have reciprocal preferences towards commercial policy since such preferences are exhibited by voters. In experiments, there is ample evidence of both positive and negative reciprocity in individual decision making. Public opinion polls suggest that a significant proportion of people do

---

2 See http://www.ustr.gov/node/4442.
6 In fact, one could argue that governments are not genuinely motivated by fairness considerations, but only push for fairer trade policies in order to increase their political support, since voters have such considerations. In any case, our results depend solely on governments exhibiting, to some degree, reciprocal preferences towards commercial policy, and not on why this is the case.
7 For instance, experiments asking individuals to contribute to public goods typically find that their contributions far exceed what self-interested utility maximization would entail (e.g., Andreoni, 1988; Palfrey and Prisbrey, 1997; Croson, 2007). This is usually interpreted as evidence of positive reciprocity. Analogous results arise from trust or gift-exchange experiments (e.g., Berg et al., 1995; Fehr et al., 1997; Fehr et al., 1998). On the other hand, evidence for negative reciprocity is found in ultimatum-game experiments with the typical result being that people reject offers that would be accepted under the self-interested hypothesis.
exhibit reciprocal preferences with respect to trade policy as well. For example, in a January 2004 PIPA poll, 67% of Americans agreed that "in general, if another country is willing to lower its barriers to products from the US if we [the US] will lower our [its] barriers to their products," the US should do so, whereas only 24% disagreed with this statement. More importantly, almost 75% of the ones endorsing this statement agreed that "the US should only lower its barriers if other countries do, because that is the only way to pressure them to open their markets," while just 24% of them thought that "the US should lower its barriers even if other countries do not, because consumers can buy cheaper imports and foreign competition spurs American companies to be more efficient." Similar preferences towards commercial policy are exhibited by Europeans. In a spring 2001 Eurobarometer survey, more than 74% of EU-15 citizens did endorse reciprocity in international trade agreements, whereas merely 7% of them did not. Another important conclusion that could be drawn from these polls is that people have an explicit notion of what fair trade involves, i.e., they have a reference level regarding fair commercial policy against which the policies implemented by their own government or by the rest of the world can be evaluated. For instance, in a July 2004 CCFR poll, Americans were asked whether the US practices fair trade with various other countries, and whether the countries in question have fair trade policies towards the US. The percentage of respondents who were not sure/declined to answer was very low overall, ranging from 11% to 14% in the former case and from 11% to 15% in the latter. Likewise, in a 19-nation poll conducted November 2003 through February 2004, people around the globe were asked whether rich countries are playing fair in trade negotiations with poor countries. In this poll, the percentage of respondents who

(e.g., Güth et al., 1982; Roth et al., 1991). Moreover, in a recent paper, Dohmen et al. (2009) provide evidence of both positive and negative reciprocity using survey data.


10The "reference level" is a concept widely studied in the behavioral economics literature. For more on this, see, for instance, Helson (1964) and Tversky and Kahneman (1991).

were not sure/declined to answer was just 10% or lower in 14 out of the 19 countries.\textsuperscript{12}

On a more theoretical level, if individuals have reciprocal preferences, political-economy models of trade policy suggest that these preferences will be reflected in the government’s objective function. The first model we can invoke here is the median-voter model, where the government chooses policies that reflect majority opinion on the issue (in order to remain popular and stay elected).\textsuperscript{13} In such a setting, if the median voter has reciprocal preferences, then the government’s actions are going to mirror these preferences. Instead, we could look at interest-group models. The framework that currently occupies center stage in the literature is due to Grossman and Helpman (1994). In their paper, the incumbent government maximizes a weighted average of aggregate social welfare and political contributions by lobbies that wish to influence trade policy. Alternatively, political influences could be readily represented, as Baldwin (1987) has demonstrated, by a parameter that attaches additional weight to producer surplus in the government’s objective function. In either case, if individuals have preferences for fairness and reciprocity, these preferences will enter into the government’s objective function with some weight.\textsuperscript{14}

To address the implications of reciprocity and fairness for commercial policy, we develop a dynamic game in which reciprocal countries facing a terms-of-trade Prisoner’s Dilemma problem in their dealings with one another attempt to maintain tariff cooperation. To model reciprocity, we follow Segal and Sobel (2007). In particular, we assume that a country attaches a positive (negative) weight to the self-interested welfare of a trading partner if it expects the latter to behave kindly (unkindly) by imposing an import tariff that is lower (higher) than the one it perceives as fair. In other words, countries are assumed to have preferences over both outcomes and strategies. Of course, fairer tariff policies correspond to a fairer distribution of the gains from trade among countries (or interest

\textsuperscript{12}See http://www.worldpublicopinion.org/pipa/pdf/jun03/GlobalIss__Jun04__quaire.pdf.
\textsuperscript{13}See, for example, Mayer (1984).
\textsuperscript{14}See Goldberg and Maggi (1999) and Gawande and Bandyopadhyay (2000) for US empirical evidence that social welfare does in fact receive a high weight in the government’s objective function.
groups in different countries), meaning that our results could be readily reinterpreted in terms of the distribution of the trade gains among trading partners. Besides the countries’ reciprocal preferences, we model trade agreements in a standard fashion. More specifically, we maintain the assumption that binding commitments cannot be made at the international level and countries are therefore limited to cooperative multilateral tariff agreements that are self-enforcing. This assumption reflects the lack of a strong mechanism within the WTO for enforcing the trade policies agreed upon under its auspices. In this context, a country will choose today to adhere to the cooperative path as long as the onetime gain it could achieve by unilaterally deviating from its agreed-upon trade policies does not outweigh the discounted welfare cost of the future trade war its defection would ignite.\textsuperscript{15}

We find that reciprocal countries that are moderately demanding from their trading partners with respect to their commercial policy (i.e., when the tariffs considered fair are not too low, but at the same time, overly restrictive import policies are perceived as unfair) can support a greater degree of multilateral tariff cooperation and thus achieve higher welfare than self-interested ones. The intuition is straightforward. For such fair-tariff perceptions, in the reciprocal game (i) the punitive Nash tariffs are higher than in the self-interested one; and (ii) the countries are in a positive-reciprocity state. As a result, under the scenario in question, reciprocal countries face both a weaker incentive to defect and a stronger incentive to cooperate than self-interested ones, allowing them to maintain more liberal trade policies.

However, when reciprocal countries are highly demanding from their trading partners regarding their import policy (i.e., when only very liberal import policies are considered fair), then the effect of reciprocity on multilateral trade cooperation is ambiguous. Intuitively, in such a case, reciprocal countries are in a negative-reciprocity state, meaning that they face a stronger incentive to cheat than self-interested ones, even though their incentive \textsuperscript{15} See, for example, Dam (1970), Dixit (1987), and Bagwell and Staiger (2002) for further elaboration on these points.
to cooperate is still relatively stronger due to the harsher punishment a defector faces in the reciprocal game. Our simulations do confirm that for very low fair tariffs, there are indeed cases where self-interested countries can support lower cooperative tariffs in equilibrium than reciprocal ones. Finally, we demonstrate that our results are robust to allowing for fair-tariff perceptions that are endogenously determined during the course of the game.

At this point, it is important to note that our findings suggest a novel reason for the occasional failure of trade negotiations: Assuming countries have (some) preferences for fairness and reciprocity, if they arrive at the negotiating table with expectations that are highly elevated (i.e., they have very low fair-tariff perceptions), this could prove counter-productive, in the sense that they might no longer be able to support very liberal trade policies. As a matter of fact, Mr. Renato Ruggiero, the WTO Director-General in 1995-1999, referred to this possibility in one of his speeches in 1995: "I have heard it said that unrealistically high expectations could pose a threat to the success of the negotiations. I have also heard it suggested that failure to meet such expectations could make a multilateral solution impossible."\(^{16}\) It could then be argued that this might be one of the possible explanations for the problems plaguing the Doha Round since its launch in 2001. More specifically, the success of the 1986-94 Uruguay Round as well as the deepening of globalization in general in the 1990s might have led to countries entering the Doha negotiations with expectations that were too high, hindering the efforts for further multilateral trade liberalization. Put differently, if countries had entered the Doha Round with lower expectations, its outcome might have been more favorable. Another plausible explanation, still along the lines of our model, is that developing countries might have arrived at the negotiations being too demanding from developed countries regarding their trade policies (especially with respect to agricultural goods), partly due to their feeling that the previous round was lopsided or unfair.\(^{17}\) In any case, as Dani Rodrik writes: "In the end, it may

\(^{16}\)See http://www.wto.org/english/news_e/pres95_e/pr9512_e.htm.

\(^{17}\)For more on the latter explanation, see "The Doha Round...and Round...and Round..." The Economist (print edition), July 31, 2008.
well be the atmospherics — psychology and expectations — rather than the actual economic results on the ground that will determine the outcomes [of the Doha Round]." In summary, expectations emerge as a key factor in our analysis, having a significant effect on what can be achieved in the international trade arena. In fact, to the best of our knowledge, our paper is the first to identify such a role for expectations in international trade negotiations. The policy implications of our model are then straightforward. The careful management of expectations is critical for the success of multilateral trade negotiations, and most importantly, the creation of a pre-negotiations high-expectations environment should be avoided.

The remainder of the paper is organized as follows. Section 2 sets out the basics. Section 3 characterizes the static Nash equilibrium of our model, whereas Section 4 analyzes the dynamic game. Section 5 presents a simplified model in order to better illustrate the main insights from our analysis. Section 6 endogenizes countries' fair-tariff perceptions. Finally, Section 7 identifies some promising avenues for future research and concludes. All the proofs are relegated to the Appendix.

2 The Model

We assume the world consists of two countries, A and B, that trade two goods, a and b. This focus on a 2-country world is nonrestrictive, as our findings readily extend to N countries. Country J is endowed with 1 unit of good \(-j\) and zero units of good \(j\), where \(J \in \{A, B\}\) and \(j \in \{a, b\}\). On the consumption side, we maintain the assumptions that demand functions are symmetric across countries and goods, and that the demand for good \(j\) is independent of the price of good \(-j\). More specifically, the demand for good \(j\)

---

18 See http://www.guardian.co.uk/commentisfree/2008/aug/08/wto.internationalaidanddevelopment.
19 Our framework is inspired by Bagwell and Staiger (1999b).
20 Upon request, a technical appendix is available from the authors in which an N-country model is presented.
21 We choose to ignore the production process in the two countries for expositional simplicity. In any case, this assumption does not affect the qualitative nature of our findings.
in country $J$ is given by $D \left( P^j_J \right)$, where $P^j_J$ is good $j$’s price in country $J$. We make the standard assumptions that $D \left( P^j_J \right)$ is strictly positive on some bounded interval $[0, \overline{P}^j_J]$, that $D \left( P^j_J \right) = 0$ for $P^j_J \geq \overline{P}^j_J$, and that $D \left( P^j_J \right)$ is twice continuously differentiable in $P^j_J$ with $D' \left( P^j_J \right) < 0$ for $P^j_J \in [0, \overline{P}^j_J]$. Given our setup, country $J$ imports good $j$, whereas it exports good $-j$ in accordance with the following export supply function:

$$X^{-j}_J (P^{-j}_J) = 1 - D \left( P^{-j}_J \right).$$  \hspace{1cm} (1)

In each period, the countries simultaneously select specific (nonprohibitive) import tariffs so as to maximize their individual welfare. The tariffs are picked with perfect information as to all past tariff choices. Let country $J$’s import tariff be $\tau^J \in \Theta^J \subset \mathbb{R}^+$, where $\Theta^J$ is a compact interval. The no-arbitrage condition for good $j$ yields:

$$P^j_J = P^{-j}_J + \tau^J.$$  \hspace{1cm} (2)

The equilibrium prices can then be obtained from the usual market-clearing conditions:

$$D \left( P^j_J (\tau^J) \right) = X^{-j}_J \left( P^{-j}_J (\tau^J) \right).$$  \hspace{1cm} (3)

The countries are assumed to have preferences for fairness and reciprocity. In particular, the welfare of country $J$ is given by:

$$RW^J \left( \tau^J, \tau^{-J}, \tau^{-j}_J \right) = SW^J \left( \tau^J, \tau^{-J} \right) + \gamma \omega^j \left( \tau^{-J}, \tau^{-j}_J \right) SW^{-j} \left( \tau^J, \tau^{-J} \right).$$  \hspace{1cm} (4)

The first term, $SW^J$, is the self-interested (or "standard") welfare function, i.e., the sum

---

22 An alternative way of modeling reciprocity in a dynamic setup is due to Dufwenberg and Kirchsteiger (2004). In their paper, they develop a theory of reciprocity for extensive form games, and introduce a new solution concept – sequential reciprocity equilibrium – where players update their beliefs about their co-players’ intentions as the game unfolds and choose their actions accordingly. However, their framework would be highly intractable for the purposes of this paper.
of consumer surplus, producer surplus, and tariff revenue:

\[ SW^J (\tau^J, \tau^{-J}) = \int_{P^J_j(\tau^{-J})}^{P^J_j} D(P) \, dP + \int_{P^J_j(\tau^{-J})}^{P^J_j} D(P) \, dP + P^J_j(\tau^{-J}) + \tau^J X^{-J}_j (\tau^J) . \]  

(5)

The second term, \( \gamma w^J(\tau^{-J}, \tau_f^{-J}) SW^{-J} \), captures the fairness payoff for country \( J \), where (i) its relative significance is specified by the scaling factor \( \gamma \geq 0 \); and (ii) \( w^J(\tau^{-J}, \tau_f^{-J}) \) determines the weight country \( J \) places on its trading partner’s self-interested welfare \( SW^{-J} \), and is of the following form:

\[
\begin{cases}
  > 0 & \text{if } \tau_f^{-J} > \tau^{-J} \\
  = 0 & \text{if } \tau^{-J} = \tau_f^{-J} \\
  < 0 & \text{otherwise}
\end{cases}
\]

(6)

with \( \tau_f^{-J} \) being the \( \tau^{-J} \) country \( J \) deems "fair." We maintain the assumptions that country \( J \)’s weight function \( w^J(\tau^{-J}, \tau_f^{-J}) \) is twice continuously differentiable in both arguments, and is nondecreasing in its own fair-tariff perception and nonincreasing in country \(-J\)’s tariff. Furthermore, we assume that fair-tariff perceptions are common knowledge.

Intuitively, the function \( w^J \) reflects the fact that a reciprocal country cares about the intentions of its trading partner. More specifically, the first condition in (6) expresses positive reciprocity: If country \( J \) expects the tariff of country \(-J\) to be smaller than its own perception of a fair tariff, then it is willing to sacrifice some of its self-interested welfare to reward its trading partner. On the other hand, the third condition in (6) expresses negative reciprocity: When country \( J \) expects country \(-J\) to impose a tariff that exceeds the one it perceives as fair, then it wishes to punish its trading partner and is willing to sacrifice some of its own self-interested welfare in order to do so. Moreover, if \( \tau^{-J} \) and \( \tau_f^{-J} \) are equal, then the self-interested and the reciprocal welfare functions coincide for country \( J \). In brief, equation (6) signifies that from country \( J \)’s standpoint, any \( \tau^{-J} \) below \( \tau_f^{-J} \) is a fair (or kind) action on the part of country \(-J\) that should be rewarded, whereas any
\(\tau^{-J}\) in excess of \(\tau_f^{-J}\) is an unfair (or unkind) action that should be punished.

3 Static Game

Our aim in this section is to characterize the static Nash equilibrium of our model, and compare it with the one that would emerge in a game with self-interested countries. This equilibrium will serve as a credible punishment in the dynamic game considered in the next section, the threat of which can support multilateral tariff cooperation in a repeated setting.\(^{23}\) To this end, let the static game with self-interested countries be denoted by \(\Gamma^S(SW)\), while \(\Gamma^R(RW, w, \tau_f)\) represents the static game with reciprocal countries, where \(\tau_f \equiv (\tau_f', \tau_f^-)\) is the fair-tariff vector. We henceforth assume that \(\tau_f' = \tau_f^- \equiv \tau_f\), i.e., the countries have a common fair-tariff perception. The reason for this assumption is twofold. First, it considerably simplifies the analysis. Second, asymmetries of a not too high degree in fair-tariff perceptions between the (otherwise symmetric) countries would not affect the qualitative nature of our findings.\(^{24}\)

It is direct to show that the cross-partial derivative of the welfare function of reciprocal country \(J\) with respect to its own tariff and country \(-J\)'s tariff is nonnegative (i.e., \(\frac{\partial RW_J}{\partial \tau J \partial \tau^{-J}} \geq 0\)). This means country \(J\)'s incremental returns from raising its own tariff are nondecreasing in its partner's tariff, i.e., the choice variables are strategic complements. On the other hand, the cross-partial derivative of country \(J\)'s welfare function with respect to its tariff and its fair-tariff perception is nonpositive (i.e., \(\frac{\partial RW_J}{\partial \tau J \partial \tau_f} \leq 0\)). To gain some insight for the sign of these derivatives, simply recall that (i) increasing \(\tau_f\) inflicts a negative terms-of-trade externality on country \(-J\) (given our countries are "large"); and (ii) a higher \(\tau_f^- (\tau_f)\) results ceteris paribus in a smaller (larger) \(w^J\).

\(^{23}\) In fact, the static Nash equilibrium would be the unique equilibrium for the dynamic game as well if a multilateral trade agreement were not feasible (e.g., due to exogenous, political reasons or because the countries were highly impatient and did not value the future at all).

\(^{24}\) Upon request, a technical appendix is available from the authors in which we reproduce our analysis allowing for asymmetries in fair-tariff perceptions between the countries that are not too high. See also footnote 29.
Our first result establishes the existence of a pure symmetric Nash equilibrium for both $\Gamma^R(RW, w, \tau_f)$ and $\Gamma^S(SW)$.

**Lemma 1** For the static game with reciprocal countries $\Gamma^R(RW, w, \tau_f)$, there exist largest and smallest pure symmetric Nash equilibria, $\tau_{NR}^* \equiv (\tau_{NR}, \tau_{NR})$ and $\tau_{NR}^* \equiv (\Sigma_{NR}, \Sigma_{NR})$. Moreover, for the static game with self-interested countries $\Gamma^S(SW)$, there also exist largest and smallest pure symmetric Nash equilibria, $\tau_{NS}^* \equiv (\tau_{NS}, \tau_{NS})$ and $\tau_{NS}^* \equiv (\Sigma_{NS}, \Sigma_{NS})$.

We next show how countries’ fair-tariff perception affects the extremal equilibrium tariffs of $\Gamma^R(RW, w, \tau_f)$.

**Lemma 2** The largest and the smallest pure Nash equilibria of $\Gamma^R(RW, w, \tau_f)$, i.e., $\tau_{NR}^* \equiv (\tau_{NR}, \tau_{NR})$ and $\tau_{NS}^* \equiv (\Sigma_{NR}, \Sigma_{NR})$, are nonincreasing in $\tau_f$.

Intuitively, as we argued above, for a given $\tau^{-J}$, a higher $\tau_f$ leads to a larger $w^J$. Consequently, country $J$ now wishes to reduce the terms-of-trade negative externality of its tariff on its trading partner when choosing its import policy, resulting in more liberal Nash tariff equilibria.

It turns out that all our results hold independently of whether we consider the largest or the smallest pure Nash equilibria of $\Gamma^R(RW, w, \tau_f)$ and $\Gamma^S(SW)$. Therefore, without loss of generality, we drop from now on the "bar" notation and simply write $\tau_{NR}^* \equiv (\tau_{NR}, \tau_{NR})$ and $\tau_{NS}^* \equiv (\tau_{NS}, \tau_{NS})$, referring to either $\tau_{NR}^*$ and $\tau_{NS}^*$, or $\tau_{NR}^*$ and $\tau_{NS}^*$, respectively. In addition, we hereafter assume that (i) $\gamma \neq 0$ so that country $J$’s fairness payoff is nonzero for $\tau^{-J} \neq \tau_f$; and (ii) $\tau_f \leq \tau_{NS}$, i.e., overly restrictive import policies are considered unfair, which is a reasonable assumption given our focus on trade cooperation among countries.

We next compare $\tau_{NR}$ with $\tau_{NS}$ as well as the welfare obtained in $\Gamma^R(RW, w, \tau_f)$ and $\Gamma^S(SW)$.

**Proposition 1** Under our model’s assumptions, (i) $\tau_{NR} \geq \tau_{NS}$; and (ii) for any $J$, $RW^J(\tau_{NR}, \tau_f) \leq SW^J(\tau_{NS})$, with equality only holding for $\tau_f = \tau_{NS}$.  

12
Proposition 1 demonstrates that reciprocal countries end up with a more protectionist and thus welfare-inferior Nash equilibrium than self-interested countries. The intuition is straightforward. At $\tau_{NS}$, reciprocal countries are in a negative-reciprocity state wishing to punish each other (since $\tau_f \leq \tau_{NS}$), implying that the Nash equilibrium tariff of $\Gamma^R(RW, w, \tau_f)$ must exceed the one of $\Gamma^S(SW)$.\(^{25}\)

4 Dynamic Game

We now study repeated interaction between the countries. In particular, the dynamic game we consider is the stage game analyzed above infinitely repeated. We assume that countries cannot make binding international commitments but are instead limited to self-enforcing trade agreements. In such a setting, countries can still maintain multilateral trade cooperation, whose degree depends critically on how severely they can credibly punish an offender. Our aim in this section is to evaluate the effect of fairness and reciprocity on the ability of countries to cooperate with low import tariffs.

To this end, denote the infinitely repeated game with reciprocal countries by $\Gamma^R_\infty(RW, w, \tau_f)$, and the one with self-interested countries by $\Gamma^S_\infty(SW)$. The discount factor between periods is $\delta \in (0, 1)$. For both games, we focus on symmetric cooperative subgame-perfect equilibria in which (i) along the equilibrium path, the countries set a common cooperative tariff $\tau_C < \tau_{NS}$ in each period; and (ii) if at any point in the game a defection occurs, both countries revert from the following period onwards to the noncooperative Nash tariff of the (relevant) stage game.\(^{26}\) In other words, to enforce cooperation,\(^{25}\)

\(^{25}\)At this point, a technical note is in order. Our results do not require differentiability of $RW^f(\bullet)$. Rather, our findings would still hold under the significantly weaker assumptions that $RW^f(\bullet)$ has increasing differences in $J^\bullet; J^\tau$ and decreasing differences in $J^\tau; f$. Nevertheless, for expositional simplicity, we have chosen to work with differentiable welfare functions.

\(^{26}\)Given the overall symmetry of our framework, it is only natural to focus on symmetric equilibria, which imply an equal split between the countries of the gains from cooperation. Actually, it can be readily shown that such a split supports the highest degree of multilateral trade cooperation in our setting. On a different note, observe that for both games we restrict our attention to cooperative tariffs lower than $\tau_{NS}$. This enables us to better compare $\Gamma^R_\infty(RW, w, \tau_f)$ with $\Gamma^S_\infty(SW)$, which is our main goal in this paper.
the countries employ a grim-trigger strategy.

We begin our analysis with $\Gamma^S_{\infty} (SW)$. The static incentive self-interested country $J$ has to cheat is defined as:

$$SW^J (BR^J_S (\tau_C), \tau_C) - SW^J (\bar{\tau}^C) \equiv SW^J_D - SW^J_C \equiv \Omega^J_S (\tau_C),$$  \hspace{1cm} (7)

where $BR^J_S (\tau_C)$ is country $J$'s best-response tariff to $\tau_C$ and $\bar{\tau}^C \equiv (\tau_C, \tau_C)$. $\Omega^J_S$ equals simply the onetime increase in welfare country $J$ achieves when it optimally chooses a tariff on its reaction curve while its trading partner still cooperates with $\tau_C$. On the other hand, violating multilateral cooperation also bears consequences as a trade war ensues. The discounted future welfare cost a defector faces equals:

$$\frac{\delta}{1 - \delta} (SW^J (\bar{\tau}^C) - SW^J (\bar{\tau}^N_S)) \equiv \frac{\delta}{1 - \delta} (SW^J_C - SW^J_N) \equiv \frac{\delta}{1 - \delta} \omega^J_S (\tau_C),$$  \hspace{1cm} (8)

where $\omega^J_S$ is the per-period value of cooperation for country $J$, i.e., the per-period increase in country $J$'s welfare under multilateral cooperation as compared with its welfare during a tariff war. Therefore, the incentive-compatibility condition for a self-interested country $J$ to adhere to the cooperative path in $\Gamma^S_{\infty} (SW)$ is that the onetime gain from defection, $\Omega^J_S$, does not outweigh the discounted value of future cooperation, $\frac{\delta}{1 - \delta} \omega^J_S$:

$$\Omega^J_S (\tau_C) \leq \frac{\delta}{1 - \delta} \omega^J_S (\tau_C).$$  \hspace{1cm} (9)

From (9), it is direct to show that a given cooperative tariff $\tau_C$ can be supported as a subgame-perfect equilibrium of the dynamic game as long as countries are patient enough, or:

$$\delta \geq \delta^S_{\tau_C} \equiv \frac{SW^J_D - SW^J_C}{SW^J_D - SW^J_N}.$$  \hspace{1cm} (10)

Analogous relationships hold for countries with reciprocal preferences. In particular,
the incentive-compatibility condition for a reciprocal country $J$ to uphold multilateral cooperation is given by:

$$\Omega^J_R(\tau_C) \leq \frac{\delta}{1-\delta} \omega^J_R(\tau_C).$$

(11)

Moreover, for a given cooperative tariff $\tau_C$, the minimum discount factor required so that the tariff in question can be multilaterally sustained equals:

$$\delta^R_{\tau_C} \equiv \frac{RW^J_B - RW^J_C}{RW^J_B - RW^J_N}.$$ 

(12)

Our next lemma demonstrates that reciprocal countries can support a fairly liberal cooperative tariff as long as they are sufficiently patient.

**Lemma 3** Let $\tau_C \leq \tau_f$ be a cooperative tariff. Then a sufficiently high discount factor exists such that $\tau_C$ is a subgame-perfect Nash equilibrium for $\Gamma^R_{\infty}(RW, w, \tau_f)$.

Let us, in all that follows, maintain the (nonrestrictive) assumption that $\gamma$ is sufficiently small, meaning that the relative weight of the fairness payoff in the countries’ objective function (or, equivalently, the relative weight the countries place on their trading partner’s self-interested welfare) is not too high.\(^{27}\) We are at this point ready to state our first result about the impact of fairness and reciprocity on multilateral trade cooperation. Using (10) and (12), we now compare $\delta^S_{\tau_C}$ against $\delta^R_{\tau_C}$, where $\tau_C \leq \tau_f$. Remember that $\Gamma^R_{\infty}(RW, w, \tau_f)$ and $\Gamma^S_{\infty}(SW)$ are identical in all respects except for the fairness payoff in the countries’ welfare function.

**Proposition 2** Let $\tau_C \leq \tau_f$ be a cooperative tariff. The critical discount factor above which multilateral cooperation can be maintained at $\tau_C$ is lower in $\Gamma^R_{\infty}(RW, w, \tau_f)$ than in $\Gamma^S_{\infty}(SW)$, i.e., $\delta^R_{\tau_C} < \delta^S_{\tau_C}$.

\(^{27}\)The derivation of a closed-form solution for the upper bound of $\gamma$ has proved elusive. In the simulations in the next section where a simplified model is presented, $\gamma$ is set less than or equal to 0.1.
The intuition underlying Proposition 2 is straightforward once it is recalled that for any cooperative tariff \( \tau_C \) lower than the fair tariff \( \tau_f \), the countries attach a positive weight to their partner’s self-interested welfare, i.e., they are in a positive-reciprocity state. Two reinforcing forces are at work here. On the one hand, for any country \( J \), the value of cooperation at \( \tau_C \) is higher in \( \Gamma^R_\infty(RW, w, \tau_f) \) than in \( \Gamma^S_\infty(SW) \) since in the former game (i) the noncooperative (punitive) Nash tariff is higher; and (ii) infinite Nash reversion would also be costly for \( J \)’s trading partner, which acts to heighten the cost of the punishment phase for country \( J \) itself. On the other hand, the static incentive country \( J \) has to deviate from \( \tau_C \) is weaker in \( \Gamma^R_\infty(RW, w, \tau_f) \) than in \( \Gamma^S_\infty(SW) \) because in the former game (i) the defect tariff is lower, since the countries are in a positive-reciprocity state; and (ii) defection would hurt \( J \)’s partner, mitigating \( J \)’s potential onetime gains from cheating. It then follows that reciprocal countries can more easily support any given cooperative tariff below the fair tariff than self-interested ones.

Let now \( \tau_{CS} \equiv (\tau_{CS}, \tau_{CS}) \) be the most cooperative equilibrium tariff vector of \( \Gamma^S_\infty(SW) \), i.e., \( \tau_{CS} \) is the smallest nonnegative tariff that does not invite cheating in the dynamic game with self-interested countries. Similarly, \( \tau_{CR} \equiv (\tau_{CR}, \tau_{CR}) \) is the most cooperative equilibrium tariff vector of \( \Gamma^R_\infty(RW, w, \tau_f) \).\(^{28}\) Clearly, \( \tau_{CS} \) (\( \tau_{CR} \)) is the most cooperative equilibrium tariff of \( \Gamma^S_\infty(SW) \) (\( \Gamma^R_\infty(RW, w, \tau_f) \)) when \( \delta = \delta_{\tau_{CS}}^S \) (\( \delta = \delta_{\tau_{CR}}^R \)). Moreover, we assume in the remainder of this section that \( \delta \in [\delta, \delta] \) so that both self-interested and reciprocal countries can maintain some cooperation in trade policies but global free trade is infeasible for either of them. The next proposition compares \( \tau_{CR} \) with \( \tau_{CS} \) assuming the countries are moderately demanding from their trading partners regarding their commercial policy (i.e., assuming the fair tariff is not too low).

**Proposition 3** Let \( \tau_f \geq \tau_{CS} \). Then the most cooperative equilibrium tariff of \( \Gamma^S_\infty(SW) \) is higher than the one of \( \Gamma^R_\infty(RW, w, \tau_f) \), i.e., \( \tau_{CS} > \tau_{CR} \).

\(^{28}\) Note that the most cooperative equilibrium is the most natural focal point of either game since (i) it is the only equilibrium of the desired class that is not Pareto dominated; and (ii) nothing precludes preplay communication between the countries.
The intuition underlying Proposition 3 is the same as the one behind Proposition 2.

Finally, we compare $\tau_{CR}$ with $\tau_{CS}$ assuming now that the countries are highly demanding from their trading partners with respect to their import policy (i.e., assuming that only very liberal import policies are considered fair).

**Proposition 4** Let $\tau_f < \tau_{CS}$. Then the effect of fairness and reciprocity on the most cooperative tariff equilibrium of the dynamic game is ambiguous.

To gain some insight for Proposition 4, recall that for any cooperative tariff $\tau_C$ higher than the fair tariff $\tau_f$, the countries attach a negative weight to their partner’s self-interested welfare, i.e., they are in a negative-reciprocity state. Two observations can then be readily made for any such $\tau_C > \tau_f$. On the one hand, for any country $J$, the value of cooperation at $\tau_C$ is higher in $\Gamma^R_{\infty}(RW, w, \tau_f)$ than in $\Gamma^S_{\infty}(SW)$ since the punitive Nash tariff is higher in the former game. Of course, infinite Nash reversion would be costly also for country $-J$, which acts to lower the cost of the punishment phase for country $J$ in $\Gamma^R_{\infty}(RW, w, \tau_f)$, but this effect is relatively weak for a sufficiently small $\gamma$. On the other hand, country $J$ has a stronger incentive to deviate from $\tau_C$ in $\Gamma^R_{\infty}(RW, w, \tau_f)$ than in $\Gamma^S_{\infty}(SW)$ since in the former game (i) the defect tariff is higher, because the countries are in a negative-reciprocity state; and (ii) defection would hurt country $-J$, raising the gains from cheating for country $J$. Hence, it is ambiguous whether reciprocal or self-interested countries can more easily sustain any given cooperative tariff above the fair tariff. As a result, when only very liberal trade policies are considered fair, the overall effect of reciprocity on multilateral tariff cooperation could be negative. This is more clearly illustrated in the next section within the context of a simplified model.\(^{29}\)

At a more general level, Propositions 3 and 4 demonstrate that if, for whatever reason,\(^{29}\)

\(^{29}\)At this point, we should note that asymmetries in fair-tariff perceptions would not affect the qualitative nature of our findings as long as the countries remained *symmetrically* demanding from each other with respect to their trade policy. In particular, under asymmetric fair-tariff perceptions, Proposition 3 would still hold as long as $\tau_{J}^{-J} \geq \tau_{CS}^{-J}$ for any country $J$, whereas Proposition 4 would be still valid as long as $\tau_{J}^{-J} < \tau_{CS}^{-J}$ for all $J$.\(^{17}\)
countries become more demanding from their trading partners with respect to their import policy (i.e., if the fair tariff decreases), a given cooperative equilibrium that could have been otherwise supported, might no longer be feasible. This then suggests that if countries enter a round of multilateral trade negotiations with elevated expectations due to economic and/or political reasons, they might fail to reach an agreement on further multilateral trade liberalization, even though such an agreement might have been attainable in the absence of these high expectations. Therefore, Propositions 3 and 4 provide a novel perspective on the occasional failures of multilateral trade negotiations, as in the ongoing Doha Round.

5 Simplified Model

In this section, we reproduce the results of the paper within a simple setup with linear demand curves and a specific functional form for \( w^J(\tau^{-J}, \tau_f) \). This serves two goals. First, it enables us to better illustrate the insights from our model. At the same time, the results obtained here can be more easily related to the ones found in a substantial part of the literature on trade agreements that uses similar demand specifications (e.g., Bagwell and Staiger, 1999b; Freund, 2000; Ornelas, 2005; Saggi and Yildiz, 2010). To this end, let the demand for good \( j \) in country \( J \) be given by:

\[
D(P^J_j) = \alpha - \beta P^J_j, \tag{13}
\]

where \( \alpha > \frac{1}{2}, \beta > 0 \) are constants. Moreover, let us assume that the weight function \( w^J \) is of the following form:

\[
w^J(\tau^{-J}, \tau_f) = \frac{\tau_f - \tau^{-J}}{\tau_f + \tau^{-J}} \in (-1, 1). \tag{14}
\]

We first look at the static game, and in particular at \( \Gamma^S(SW) \). It turns out the best-
response tariff of country $J$ equals:

$$BR^J_S(\tau^{-J}) = \frac{1}{3\beta},$$

(15)

meaning both countries have the same dominant strategy. Then, trivially:

$$\tau_{NS} = \frac{1}{3\beta}.$$  

(16)

However, the analysis for $\Gamma^R(RW, w, \overline{\tau}_f)$ is slightly more involved. The countries no longer have a dominant strategy. Rather:

$$BR^J_R(\tau^{-J}) = \frac{(1 + \gamma)\tau^{-J} + (1 - \gamma)\tau_f}{\beta[3(\tau^{-J} + \tau_f) + \gamma(\tau^{-J} - \tau_f)]}.$$  

(17)

It is direct to show that $BR^J_R(\tau^{-J})$ is strictly decreasing in $\tau_f$ and strictly increasing in $\tau^{-J}$. Simple algebra then yields:

$$\tau_{NR} = \frac{1 + \gamma - \beta(3 - \gamma)\tau_f + \sqrt{4\beta(1 - \gamma)(3 + \gamma)\tau_f + (1 + \gamma - \beta(3 - \gamma)\tau_f)^2}}{2\beta(3 + \gamma)}.  

(18)

Two conclusions can be drawn from equation (18). First, if $\tau_f = \tau_{NS} = \frac{1}{3\beta}$, then $\tau_{NR}$ collapses to $\tau_{NS}$. Second, we have that $\frac{\partial \tau_{NR}}{\partial \tau_f} < 0$, implying that if $\tau_f < \tau_{NS}$, then $\tau_{NR} > \tau_{NS}$, which is in line with Proposition 1.

Next, we turn to the dynamic game. Straightforward calculations reveal that the most cooperative equilibrium tariff for $\Gamma^S_{\infty}(SW)$ equals:

$$\tau_{CS} = \frac{3 - 5\delta}{\beta(9 - 3\delta)},$$

(19)

meaning that free trade could be supported by self-interested countries for $\delta \geq 3/5$.

In order to now compare $\tau_{CS}$ with the most cooperative equilibrium tariff of $\Gamma^R_{\infty}(RW, w, \overline{\tau}_f), \tau_{CR}$, we resort for expositional simplicity to graphical/numerical analy-
sis, while maintaining the (nonrestrictive) assumptions that $\tau_f < \tau_{NS}$ and $\tau_C < \tau_{NS}$.

Let us consider first the per-period value of cooperation for country $J$ in $\Gamma^R_\infty(RW, w, \tau_f)$ and $\Gamma^S_\infty(SW)$: $\omega^R_J$ and $\omega^S_J$, respectively. Figure 1 depicts the relation between the two: $\omega^R_J > \omega^S_J$ for any $\tau_C \in (0, \tau_{NS})$. Intuitively, two forces are at work here. First, the punitive Nash tariff is higher in $\Gamma^R_\infty(RW, w, \tau_f)$ than in $\Gamma^S_\infty(SW)$, i.e., $\tau_{NR} > \tau_{NS}$. Second, infinite Nash reversion would also be costly for $J$’s partner. This could raise or lower the cost of the punishment phase for country $J$ itself in $\Gamma^R_\infty(RW, w, \tau_f)$, depending on whether the countries are in a positive- or a negative-reciprocity state, i.e., depending on whether $\tau_C$ is below or above $\tau_f$. In any case, for sufficiently low $\gamma$, this effect is relatively weak.

We next examine the static incentive country $J$ has to cheat in $\Gamma^R_\infty(RW, w, \tau_f)$ and $\Gamma^S_\infty(SW)$: $\Omega^R_J$ and $\Omega^S_J$, correspondingly. As Figure 2 reveals, the former is weaker if and only if $\tau_C \in (0, \tau_f)$. The intuition is straightforward. For any given $\tau_C$ below $\tau_f$, the countries are in a positive-reciprocity state. This has a dampening effect on the defect tariff. Moreover, defection would be costly for $J$’s partner, which acts to mitigate the potential onetime gains from cheating for $J$ in $\Gamma^R_\infty(RW, w, \tau_f)$. As a result, $\Omega^R_J < \Omega^S_J$ for any such $\tau_C$. Of course, the reverse is true for cooperative tariffs above the fair tariff, i.e., $\Omega^R_J > \Omega^S_J$ for $\tau_C > \tau_f$.

Therefore, for $\tau_f \in [\tau_{CS}, \tau_{NS})$, reciprocal countries have a stronger incentive to cooperate and a weaker incentive to defect than self-interested countries around $\tau_{CS}$, implying that the former can support more liberal trade policies than the latter, or $\tau_{CR} < \tau_{CS}$. However, for $\tau_f < \tau_{CS}$, reciprocal countries have around $\tau_{CS}$ both a stronger incentive to cheat and a stronger incentive to cooperate than self-interested ones. In other words, there are two offsetting forces at play for low fair-tariff perceptions, making the comparison between $\tau_{CS}$ and $\tau_{CR}$ less clear-cut. Our simulations do confirm that for very low fair tariffs, $\tau_{CR}$ does indeed exceed $\tau_{CS}$, as we depict in Figure 3. Actually, it is interesting to note that $\tau_{CR}$ is more likely to exceed $\tau_{CS}$ when $\delta$ is relatively low, i.e., when the countries are

---

30The analysis was carried out using Mathematica. The file is available from the authors upon request.
relatively impatient. This is due to the fact that a lower δ weakens the relative significance of the stronger-incentive-to-cooperate force, while it leaves the stronger-incentive-to-cheat force unaffected. To summarize, when countries are highly demanding from their trading partners regarding their commercial policy, reciprocity could have a detrimental effect on multilateral tariff cooperation, and this is more likely to occur if countries are relatively impatient.

6 Endogenizing Fairness

We have hitherto assumed that perceptions of fairness are exogenous and constant between periods. This is consistent with the experimental work of Fehr and Falk (1999) who in a wage-setting context find virtually no change in either behavior or perceptions of fairness over time. However, one could argue that countries’ perceptions of a fair tariff might adjust during the course of the game. As Kahneman et al. (1986, pp.730-1) write: "Psychological studies of adaptation suggest that any stable state of affairs tends to become accepted eventually, at least in the sense that alternatives to it no longer readily come to mind. Terms of exchange that are initially seen as unfair may in time acquire the status of a reference transaction...they [people] adapt their views of fairness to the norms of actual behavior."31 In this section, we extend our analysis by allowing for endogenous formation of fair-tariff perceptions, and investigate whether our main predictions so far still hold.

To this end, we adapt to our discrete-time framework an equation widely used in the habit formation literature (e.g., Ryder and Heal, 1973; Carroll et al., 2000; Fuhrer, 2000), assuming current tariffs affect future fair-tariff perceptions as follows:

\[
\tau_f^{-J,t} = \alpha \tau_f^{-J,t-1} + (1 - \alpha) \tau_f^{-J,t-1}, \text{ for any } J, \tag{20}
\]

where \(\alpha \in (0, 1)\). Equation (20) indicates that country \(J\)'s fair-tariff perception today is a

---

31 See Franciosi et al. (1995) for experimental support of these ideas in a price-setting context.
linear combination of its last period’s fair-tariff perception and of the tariff it actually faced in that period. In other words, if country \(-J\) behaves kindly today by choosing an import tariff below the one country \(J\) deems fair, then country \(J\) will be more demanding next period (i.e., \(\tau^{-J}_{f-1}\) decreases), which acts to lower its fairness payoff for any given \(\tau^{-J}_{f}\). On the other hand, if country \(-J\) imposes an unfairly high tariff today, then country \(J\) will be less demanding next period (i.e., \(\tau^{-J}_{f+1}\) increases), which acts to raise its fairness payoff given a \(\tau^{-J}_{f}\). It follows that as the game unfolds, country \(J\)’s fair-tariff perception converges to the tariff policy of its trading partner. More formally:

\[
\lim_{t \to \infty} \left| \tau^{-J}_{f-t} - \tau^{-J}_{t} \right| = 0. 
\]

We are now prepared to examine whether our main conclusions heretofore are affected in any fundamental way by (20). Let us make the assumption that \(\alpha\) is not too big (i.e., countries do not adjust their reference levels too fast). Clearly, endogenizing fairness has no impact on the static game. At the same time, in the dynamic game, the major difference equation (20) introduces is that both \(\tau_{NR}\) and \(\tau_{CR}\) vary over time. However, given an \(\alpha\) and an initial fair-tariff perception, we can readily derive the future fair-tariff perceptions, and hence \(\tau_{NR}^{t}\) and \(\tau_{CR}^{t}\) for all \(t\).

It is direct to show that if the fair tariff initially exceeds the most cooperative equilibrium tariff of the self-interested game, then \(\tau_{CR}\) remains below \(\tau_{CS}\) along the equilibrium path, which is along the lines of Proposition 3. Intuitively, under this scenario, the countries start with a \(\tau_{CR}\) below \(\tau_{CS}\). As the game progresses, \(\tau_{f}\) converges to \(\tau_{CR}\), and thus, \(\tau_{CR}\) converges to \(\tau_{CS}\) (since \(w^{F}\) converges to zero). But for a sufficiently low \(\alpha\), \(\tau_{f}\) never falls below \(\tau_{CS}\), implying \(\tau_{CR}\) never exceeds \(\tau_{CS}\) (by Proposition 3).

Moreover, a result analogous to Proposition 4 is obtained: If the fair tariff is initially

\[32\] It is only reasonable to assume that as the multilateral trading environment becomes more liberal, countries become more demanding with respect to trade policy. For example, it is logical to expect that the tariffs deemed fair nowadays are substantially lower as compared with the ones in the late 1940s when the actual tariffs in place were significantly higher than the current ones.
smaller than $\tau_{CS}$, then the effect of fairness and reciprocity on multilateral tariff cooperation is ambiguous, since the countries might start with a $\tau_{CR}$ either below or above $\tau_{CS}$. Eventually though, under this scenario as well, the reciprocal game converges to the self-interested one (in infinite time). In summary, allowing for endogenously formed fair-tariff perceptions does not affect the qualitative nature of our findings (as long as $\alpha$ is not too high).

7 Conclusions

This paper has explored the impact of fairness and reciprocity on multilateral tariff cooperation. In particular, we examined whether reciprocal countries can sustain a greater degree of cooperation than self-interested ones in the context of self-enforcing multilateral tariff agreements. This is an important question given that governments and consumers seem to exhibit reciprocal preferences towards commercial policy. In our setting, a reciprocal country is willing to reward its trading partners by imposing lower tariffs if it expects them to behave kindly by setting their tariffs below the one it perceives as fair. However, when it expects its partners to behave unkindly by choosing tariffs in excess of the one it deems fair, it wishes to punish them and is willing to sacrifice some of its own self-interested welfare in order to do so.

We have demonstrated that as compared with self-interested countries, reciprocal ones that are moderately demanding from their trading partners regarding their commercial policy (i.e., when the commercial policies deemed fair are not too liberal) can support lower cooperative tariffs and therefore achieve higher welfare in an infinitely-repeated tariff game. Nevertheless, when countries are highly demanding from their partners with respect to their import policy (i.e., when only very liberal import policies are considered fair), reciprocity could have a detrimental effect on multilateral tariff cooperation. Our findings therefore provide a novel perspective on the successes and the occasional failures of multilateral
trade negotiations, and suggest a plausible explanation for the problems plaguing the Doha Round since its initiation in 2001. In particular, countries might have entered the Doha negotiations with too high expectations (i.e., with very low far-tariff perceptions), hindering the efforts for further multilateral trade liberalization. Finally, we have argued that our results are robust to allowing for fair-tariff perceptions that are endogenously determined during the course of the game.

In concluding, a couple of remarks are in order. First, our framework here could be readily applied to other types of agreements among countries that are constrained to be self-enforcing, such as international environmental agreements. For instance, it would be interesting to investigate whether reciprocal countries can sustain a greater degree of cooperation in abatement standards than self-interested ones. Second, another promising avenue for future research would be to examine how regional trade agreements affect countries’ fair-tariff perceptions and thus their ability to multilaterally cooperate. Given the unprecedented proliferation of such arrangements in recent years, this is a particularly important question.

A Appendix

A.1 Proof of Lemma 1

We first consider $\Gamma^R(RW, w, \overline{\tau}_f)$. Given that the number of countries is finite and that for any country $J$ (i) $\Theta^J$ is a compact interval in $\mathcal{R}^+$; (ii) $RW^J$ is twice continuously differentiable on $\Theta^J$; and (iii) $\frac{\partial RW^J}{\partial \overline{\tau}_f \overline{\tau}_f} \geq 0$, we know from Theorem 4 in Milgrom and Roberts (1990) that $\Gamma^R(RW, w, \overline{\tau}_f)$ is a (smooth) supermodular game. It then follows from Theorem 5 in Milgrom and Roberts (1990) that (i) there exist largest and smallest serially undominated strategies for each country $J$, $\overline{\tau}^J$ and $\underline{\tau}^J$; and (ii) the strategy profiles $\overline{\tau} \equiv (\overline{\tau}^J, \overline{\tau}^J)$ and $\underline{\tau} \equiv (\underline{\tau}^J, \underline{\tau}^J)$ are pure Nash equilibrium profiles. Finally, given the overall symmetry of our model, we have that $\overline{\tau}^J_{NR} = \overline{\tau}^{-J}_{NR} \equiv \overline{\tau}_{NR}$ and $\underline{\tau}^J_{NR} = \underline{\tau}^{-J}_{NR} \equiv \underline{\tau}_{NR}$.
The second part of the lemma is straightforward once it is recalled that $\Gamma^S(SW)$ can be obtained from $\Gamma^R(RW, w, \overline{\tau}_{f})$ by setting $\gamma = 0$, meaning that $\Gamma^S(SW)$ is also a (smooth) supermodular game. Q.E.D.

A.2 Proof of Lemma 2

Given that (i) $\Gamma^R(RW, w, \overline{\tau}_{f})$ is a supermodular game; and (ii) $\frac{\partial RW^J}{\partial \tau^J} \leq 0$ for any $J$, the lemma follows immediately from Theorem 6 in Milgrom and Roberts (1990). Q.E.D.

A.3 Proof of Proposition 1

If $\tau_f = \tau_{NS}$, then trivially $\tau_{NR} = \tau_{NS} \equiv \tau_N$, and for any $J$, $RW^J(\overline{\tau}_N, \tau_f) = SW^J(\overline{\tau}_N)$ since $w^J(\tau_N, \tau_f) = 0$ by (6). On the other hand, if $\tau_f < \tau_{NS}$, then $\tau_{NR} \geq \tau_{NS}$ by Lemma 2. These two inequalities imply $\tau_{NR} > \tau_f$, and thus for any $J$, $w^J(\tau_{NR}, \tau_f) < 0$ from (6). Moreover, for $\tau_{NR} \geq \tau_{NS}$, $SW^J(\overline{\tau}_{NR}) \leq SW^J(\overline{\tau}_{NS})$ for all $J$. But then it follows that for all $J$, $RW^J(\overline{\tau}_{NR}, \tau_f) < SW^J(\overline{\tau}_{NS})$. Q.E.D.

A.4 Proof of Lemma 3

If $\tau_C \leq \tau_f$, we have from (6) that for all $J$, $w^J(\tau_C, \tau_f) \geq 0$, implying:

$$RW^J(\overline{\tau}_C, \tau_f) \geq SW^J(\overline{\tau}_C).$$

(22)

In addition, we know that:

$$SW^J(\overline{\tau}_C) > SW^J(\overline{\tau}_{NS}).$$

(23)

Furthermore, we have from Proposition 1 that for any $J$:

$$RW^J(\overline{\tau}_{NR}, \tau_f) \leq SW^J(\overline{\tau}_{NS}).$$

(24)
From (22), (23), and (24), we then obtain for all $J$:

$$RW^J(\overline{\tau}_C, \tau_f) > RW^J(\overline{\tau}_{NR}, \tau_f),$$

which implies by Friedman (1971) that there exists a sufficiently high discount factor such that $\overline{\tau}_C$ is a subgame-perfect Nash equilibrium for $\Gamma^R_{\infty}(RW, w, \overline{\tau}_f)$. Q.E.D.

A.5 Proof of Proposition 2

We want to show that $\tau_C \leq \tau_f$ implies that $\delta^R = \frac{RW^J_D - RW^J_C}{RW^J_D - RW^J_N} < \frac{SW^J_D - SW^J_C}{SW^J_D - SW^J_N} = \delta^S$. To do so, we will prove:

(i) If $\tau_C \leq \tau_f \Rightarrow RW^J_D - RW^J_C \leq SW^J_D - SW^J_C$ for any $J$.

(ii) If $\tau_C \leq \tau_f \Rightarrow RW^J_D - RW^J_N > SW^J_D - SW^J_N$ for any $J$.

Let us start with (i). We have that for any $J$:

$$RW^J_C = SW^J(\overline{\tau}_C) + \gamma w^J(\tau_C, \tau_f)SW^{-J}(\overline{\tau}_C)$$

and

$$RW^J_D = SW^J(BR^J_R(\tau_C), \tau_C) + \gamma w^J(\tau_C, \tau_f)SW^{-J}(BR^J_R(\tau_C), \tau_C).$$

Therefore:

$$RW^J_D - RW^J_C = SW^J(BR^J_R(\tau_C), \tau_C) - SW^J(\overline{\tau}_C)$$

$$+ \gamma w^J(\tau_C, \tau_f) (SW^{-J}(BR^J_R(\tau_C), \tau_C) - SW^{-J}(\overline{\tau}_C))$$

$$\leq SW^J(BR^J_R(\tau_C), \tau_C) - SW^J(\overline{\tau}_C) \leq SW^J(BR^J_S(\tau_C), \tau_C) - SW^J(\overline{\tau}_C) = SW^J_D - SW^J_C.$$

We know from (6) that $w^J(\tau_C, \tau_f) \geq 0$ if $\tau_C \leq \tau_f$. Furthermore, the welfare of self-interested country $-J$ is (weakly) lower when country $J$ deviates while it still cooperates than when both countries cooperate, i.e., $SW^{-J}(BR^J_R(\tau_C), \tau_C) - SW^{-J}(\overline{\tau}_C) \leq 0$. The
first inequality then follows. The second inequality stems from the fact that \( BR_S^j(\tau_C) \) is the best reply of the self-interested country \( J \). This concludes the proof of (i).

We now turn to (ii). Let us rewrite the result we want to show:

\[
\tau_C \leq \tau_f \Rightarrow (RW_D^j - SW_D^j) - (RW_N^j - SW_N^j) > 0 \text{ for any } J.
\]

By Proposition 1 we know that the Nash equilibrium tariff of \( \Gamma^S(SW) \) is (weakly) smaller than that of \( \Gamma^R(RW, w, \tau_f) \), i.e., \( \tau_{NR} \geq \tau_{NS} \). Thus, we have that \( \tau_f \leq \tau_{NS} \leq \tau_{NR} \), implying that \( w^j(\tau_{NR}, \tau_f) \leq 0 \) by (6). Therefore, the following inequality holds for any \( J \):

\[
RW_N^j = SW^j(\tau_{NR}) + \gamma w^j(\tau_{NR}, \tau_f) SW^{-j}(\tau_{NR}) \leq SW^j(\tau_{NS}) = SW_N^j.
\]

(26)

Next we will show that \( RW_D^j - SW_D^j \geq 0 \) for any \( J \). Remember that \( \gamma \) is assumed to be sufficiently small. Taking a first-order Taylor series expansion of \( RW^j(BR_R^j(\tau_C), \tau_C, \tau_f) \) around \( \gamma = 0 \), we obtain:

\[
RW^j(BR_R^j(\tau_C), \tau_C, \tau_f) \approx SW^j(BR_S^j(\tau_C), \tau_C) + \gamma w^j(\tau_C, \tau_f) SW^{-j}(BR_S^j(\tau_C), \tau_C)
\]

\[
\Leftrightarrow RW^j(BR_R^j(\tau_C), \tau_C, \tau_f) - SW^j(BR_S^j(\tau_C), \tau_C)
\]

\[
\approx \gamma w^j(\tau_C, \tau_f) SW^{-j}(BR_S^j(\tau_C), \tau_C) \geq 0 \Leftrightarrow RW_D^j - SW_D^j \geq 0.
\]

(27)

The inequality holds due to \( w^j(\tau_C, \tau_f) \geq 0 \). By assumption, we have that \( \tau_C < \tau_{NS} \), and thus \( \tau_f \) cannot be equal to both \( \tau_C \) and \( \tau_{NS} \) at the same time. Hence, at least one of the inequalities in (26) and (27) must be strict. This concludes the proof of part (ii). Therefore, by (i) and (ii), we finally have \( \delta_R^{\tau_C} < \delta_S^{\tau_C} \). Q.E.D.
A.6 Proof of Proposition 3

From Proposition 2, we know that for any cooperative tariff \( \tau_C \leq \tau_f \), \( \delta^R_{\tau_C} < \delta^S_{\tau_C} \). So, given the assumptions of Proposition 3, this is also true for the most cooperative equilibrium tariff of the repeated game with self-interested countries, \( \tau_{CS} \): \( \delta^R_{\tau_{CS}} < \delta^S_{\tau_{CS}} \). Note that both self-interested and reciprocal countries can sustain \( \tau_{CS} \) at the discount factor \( \delta^S_{\tau_{CS}} \), but only reciprocal countries can support \( \tau_{CS} \) at \( \delta^R_{\tau_{CS}} \). From (10) and (12), we have:

\[
\begin{align*}
SW^I_D - SW^I_C &= \delta^S_{\tau_{CS}} (SW^I_D - SW^I_N) \quad \text{and} \\
RW^I_D - RW^I_C &= \delta^R_{\tau_{CS}} (RW^I_D - RW^I_N).
\end{align*}
\]

Since \( \delta^R_{\tau_{CS}} < \delta^S_{\tau_{CS}} \):

\[
RW^I_D - RW^I_C < \delta^S_{\tau_{CS}} (RW^I_D - RW^I_N) \Leftrightarrow (1 - \delta^S_{\tau_{CS}}) RW^I(BR^I_R(\tau_{CS}), \tau_{CS}, \tau_f)
\]

\[
< RW^I(\overline{\tau}_{CS}, \tau_f) - \delta^S_{\tau_{CS}} RW^I(\overline{\tau}_{NR}, \tau_f), \quad (28)
\]

meaning that \( \Omega^I_R(\tau_{CS}) < \frac{\delta^S_{\tau_{CS}}}{1 - \delta^S_{\tau_{CS}}} \omega^I_R(\tau_{CS}) \), or that the incentive-compatibility condition is not binding for a reciprocal country \( J \) at the pair \( (\tau_{CS}, \delta^S_{\tau_{CS}}) \).

Note here that \( RW^J_N \) does not depend on the cooperative tariff. Moreover, for any cooperative tariff \( \tau_C \) lower than the most cooperative equilibrium tariff of \( \Gamma^*_\infty(SW) \), \( \tau_{CS} \), the welfare for reciprocal country \( J \) under defection from \( \tau_C \) is higher than the welfare under deviation from \( \tau_{CS} \):

\[
RW^I(BR^I_R(\tau_C), \tau_C, \tau_f) > RW^I(BR^I_R(\tau_{CS}), \tau_{CS}, \tau_f).
\]

At the same time, for such a \( \tau_C < \tau_{CS} \), country \( J \)'s welfare under cooperation is also higher at \( \tau_C \) than at \( \tau_{CS} \):

\[
RW^I(\overline{\tau}_C, \tau_f) > RW^I(\overline{\tau}_{CS}, \tau_f).
\]
By the continuity of $RW^J(\cdot)$, then there exists a cooperative tariff $\tau_C < \tau_{CS}$ such that (28) still holds, or $\Omega_R^J(\tau_C) < \frac{\delta^S_{\tau_{CS}}}{1-\delta^S_{\tau_{CS}}} \omega_R^J(\tau_C)$. Since the same analysis applies to any $(\tau_{CS}, \delta^S_{\tau_{CS}})$ pair for $\delta^S_{\tau_{CS}} \in [\underline{\delta}, \overline{\delta}]$, we have that for any $\delta \in [\underline{\delta}, \overline{\delta}]$, $\tau_{CS} > \tau_{CR}$. Q.E.D.

A.7 Proof of Proposition 4

Proposition 2 holds for any cooperative tariff $\tau_C$ (weakly) lower than the fair tariff $\tau_f$. However, for any $\tau_C > \tau_f$, it is ambiguous by (25) and (27) whether $\delta^R_{\tau_C}$ or $\delta^S_{\tau_C}$ is higher, since the weight function is negative at $\tau_C$. Hence, it is possible that the minimum discount factor required for countries with reciprocal preferences to sustain cooperation at $\tau_C$ is higher than that for self-interested countries, i.e., $\delta^R_{\tau_C} > \delta^S_{\tau_C}$. Let us consider this case first, and focus on the most cooperative equilibrium tariff of $\Gamma_x^\infty(SW)$, $\tau_{CS}$. Under the scenario in question, both types of countries could sustain cooperation at $\tau_{CS}$ only with a level of discount factor equal to $\delta^R_{\tau_{CS}}$ or above. Moreover, let us make the assumption that $\Omega_R^J(\cdot)$ is a strictly convex function whereas $\omega_R^J(\cdot)$ is a strictly concave one. From (12) and (10), we have:

$$RW_D^J - RW_C^J = \delta^R_{\tau_{CS}} (RW_D^J - RW_N^J) \quad \text{and} \quad RW_D^J - RW_C^J = \delta^S_{\tau_{CS}} (SW_D^J - SW_N^J).$$

Since $\delta^R_{\tau_{CS}} > \delta^S_{\tau_{CS}}$:

$$RW_D^J - RW_C^J > \delta^S_{\tau_{CS}} (RW_D^J - RW_N^J) \Leftrightarrow (1 - \delta^S_{\tau_{CS}}) RW^J(BR^I_R(\tau_{CS}), \tau_{CS}, \tau_f)$$

$$> RW^J(\tau_{CS}, \tau_f) - \delta^S_{\tau_{CS}}RW^J(\tau_{NR}, \tau_f),$$

meaning that $\Omega_R^J(\tau_{CS}) > \frac{\delta^S_{\tau_{CS}}}{1-\delta^S_{\tau_{CS}}} \omega_R^J(\tau_{CS})$, or that the incentive-compatibility condition is violated for a reciprocal country $J$ at the pair $(\tau_{CS}, \delta^S_{\tau_{CS}})$.

---

33: This assumption is clearly not restrictive given the type of result we are here after.
For any cooperative tariff $\tau_C$ higher (lower) than the most cooperative equilibrium tariff of $\Gamma^*_\infty(SW)$, $\tau_{CS}$, the onetime gain for reciprocal country $J$ under defection from $\tau_C$ is lower (higher) than the static gain under deviation from $\tau_{CS}$:

$$\Omega^J_R(\tau_C) < (>) \Omega^J_R(\tau_{CS}).$$

At the same time, for such a $\tau_C > (\leq) \tau_{CS}$, country $J$'s per-period gain from cooperation is also lower (higher) at $\tau_C$ than at $\tau_{CS}$:

$$\omega^J_R(\tau_C) < (>) \omega^J_R(\tau_{CS}).$$

Given the strict convexity of $\Omega^J_R(\bullet)$ and the strict concavity of $\omega^J_R(\bullet)$, it follows that the incentive-compatibility condition for reciprocal country $J$ can only be restored at a cooperative tariff $\widetilde{\tau}_C > \tau_{CS}$. Since the same analysis applies to any $(\tau_{CS}, \delta_{\tau_{CS}}^S)$ pair for $\delta_{\tau_{CS}}^S \in [\underline{\delta}, \overline{\delta}]$, we have that for any $\delta \in [\underline{\delta}, \overline{\delta}]$, $\tau_{CS} < \tau_{CR}$.

Nevertheless, $\delta_{\tau_{CS}}^R < \delta_{\tau_{CS}}^S$ is also possible by (25) and (27). In this case, as we showed in the proof of Proposition 3, $\tau_{CS} > \tau_{CR}$. Therefore, when $\tau_f < \tau_{CS}$, it is ambiguous whether $\tau_{CR}$ or $\tau_{CS}$ is higher due to the ambiguity of whether $\delta_{\tau_{CS}}^R$ or $\delta_{\tau_{CS}}^S$ is higher.

Q.E.D.

References


[3] Bagwell, K., Staiger, R.W., 1999b. Regionalism and multilateral tariff cooperation,


Figure 1: The Incentive to Cooperate
Figure 2: The Incentive to Deviate

Welfare

Standard
Reciprocal

τf

τNS

Tariff
Figure 3: The Most Cooperative Tariff