

## When do fairness and reciprocity hinder multilateral tariff cooperation?

Doruk İriş\*

This study explored how fairness and reciprocity affect multilateral tariff cooperation. Reciprocal countries respond to cooperative behavior with rewards (positive reciprocity) and retaliate against unkind behavior (negative reciprocity). Previous research demonstrates that when the influence of fairness and reciprocity is modest, reciprocity promotes multilateral tariff cooperation, especially when trading partners maintain moderate demands in their commercial policies. However, cooperation is undermined if only liberal import policies are perceived as fair. This study confirms that these dynamics persist even when the impact of fairness and reciprocity is strong. Second, it identifies the conditions under which fairness and reciprocity hinder multilateral tariff cooperation.

**Key Words:** Reciprocity; Fairness; Trade agreements; Repeated games.

**JEL Classification:** F13; D90; D63.

---

(접수일 : 2024. 5. 9., 수정일 : 2024. 7. 9., 게재확정일 : 2024. 7. 11.)

\* Department of Economics, Sogang University, GN622, Mapo-gu, Seoul, 04107, S. Korea. Email: dorukiris@gmail.com.

## I. Introduction

Rabin (1993), in his seminal paper, incorporates fairness and reciprocal preferences (desire to hurt people who have mean intentions and reward people who have kind intentions) into game theory and economics. There is ample experimental evidence for reciprocal preferences among individuals (Camerer, 2003; Dohmen et al., 2009; Falk et al., 2003 and 2008; Fehr et al., 1997; Fehr and Gächter, 2002). Hadjiyiannis, İriş, and Tabakis (2012a)-henceforth HIT, asks whether such preferences matter for multilateral tariff cooperation.<sup>1)</sup>

HIT investigates the implications of reciprocal preferences for commercial policy and multilateral trade agreements. They use a multilateral tariff cooperation model inspired by Bagwell and Staiger (1999) and incorporate reciprocal preferences along the lines of Segal and Sobel (2007).<sup>2)</sup> By doing so, they provide a novel perspective on the role of expectations in multilateral trade negotiations.

More specifically, HIT develops a dynamic game in which reciprocal countries facing a terms-of-trade Prisoner's Dilemma problem in which they attempt to maintain a self-enforcing multilateral tariff agreement. Thus, countries could adhere to an agreed upon cooperative tariff if, for each country, the potential one-time gains from unilateral defection is lower than the discounted foregone future welfare due to the defection. Moreover,

---

1) Notice that reciprocal preferences are different than reciprocal strategies, commonly used in trade literature referring to mutual changes in trade policy by *self-interested* countries, such as tit-for-tat. To differentiate the two, Sobel (2005) calls the former one "intrinsic reciprocity"; we call it reciprocity or reciprocal preferences in this paper.

2) See HIT for why fairness and reciprocity might play a role in multilateral trade agreements.

reciprocal countries' welfares consist of the standard (self-interested) welfare and a fairness payoff, which captures reciprocity. If a reciprocal country expects a given trading partner to behave kindly by imposing an import tariff that is lower than its critical fair tariff perception (a reference level), then it places a positive weight to the self-interested welfare of that trading partner. However, if the given trading partner is expected to behave unkindly by setting an unfairly high tariff, then it places a negative weight on its trading partner's self-interested welfare. Thus, countries are assumed to have preferences over both outcomes and strategies, as by the latter one, countries perceive the intentions of their trading partners.

They find that reciprocal countries can support a greater degree of multilateral tariff cooperation when they are moderately demanding from their trading partners (i.e., when the tariffs considered fair are not too low.) For such fair-tariff perceptions, (i) when countries set Nash tariffs, they are in a negative-reciprocity state and, thus, the punitive Nash tariffs in the reciprocal game are higher than in the self-interested one; and (ii) when countries set cooperative tariffs, they are in a positive-reciprocity state. Both of these effects facilitates sustaining cooperation for reciprocal countries in comparison to self-interested countries. 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. For such fair-tariff perceptions, countries are in negative-reciprocity states when they either impose Nash or cooperative tariffs, yielding lower welfare levels in both cases due to negative fairness payoffs. Thus, the impact of fairness and reciprocity on the sustainability of multilateral tariff cooperation becomes ambiguous.

While HIT is the only paper, to the best of our knowledge, incorporates reciprocal preferences into international trade agreements, Lü et al. (2012) employs inequity aversion (desire to lower unfair outcome distributions) to explain the tendency of lower-earning and less-skilled intensive industries to receive relatively high levels of trade protection. On the other hand, other-regarding preferences have received more emphasis by the international environmental agreements literature. Lange and Vogt (2003) and Lange (2006) investigate the role inequality aversion (desire to lower unequal outcome distributions) in various games, such as prisoner's dilemma, symmetric emission game, and coalition formation games. Similar to HIT, Hadjiyiannis, et al. (2012b) investigates the role of fairness and reciprocity in a dynamic abatement game.<sup>3)</sup> More recently, Nyborg (2018) and İriş and Kim (2024) study the role of reciprocity in a coalition formation game and found that reciprocity could increase the coalition size.

The aim of this paper is to improve our understanding of the role of fairness and reciprocity on multilateral tariff cooperation. Thus, our contribution is twofold. First, while HIT assume the relative weight of the fairness payoff in the countries' objective function to be sufficiently small, we relax this assumption and show that all their results would still hold when the relative weight of the fairness payoff in the countries' objective function is as high as the standard welfare. This is an important extension because many experiments (Kocher et al., 2008; Fehr et al., 2014; Larney et al., 2019) find little to no effect of stake size in a number of games and the persistence of reciprocal concerns.

Second, we provide two conditions that would suffice for the impact of fairness and reciprocity to be detrimental on multilateral

---

3) See İriş and Tavoni (2020) for a review of other-regarding preferences in international environmental agreements. See also İriş and Santos-Pinto (2013) and İriş(2015, 2018) for the role of reciprocity in firm collusion.

tariff cooperation, when reciprocal countries are highly demanding from their trading partners. These conditions are: (i) the relative weight of the fairness payoff in the countries' objective function is sufficiently small, and (ii) the negative weights a country places on its trading partner's standard welfare at Nash and cooperative tariffs are sufficiently close. Notice that HIT has already showed that the incentive to cheat (i.e., one-time gain from deviation) is higher for reciprocal countries than self-interested countries under (i). On the other hand, I find that the value of cooperation is also lower for reciprocal countries than self-interested countries under (i) and (ii). The intuition behind the latter one is, under the assumption (ii) a reciprocal country would consider trading partner's cooperative tariff almost as unkind as its Nash tariff. Thus, it would weaken the value of cooperation for the reciprocal countries as similar negative fairness payoff enters to the reciprocal welfare at Nash and cooperative tariffs.

Therefore, we can tune their conclusion: if countries have some concerns for fairness and reciprocity, have high expectations from their trading partners regarding their trade policies, and also find their trading partner's cooperative tariffs almost as unkind as the non-cooperative Nash tariffs, then they will end up with a more protectionist agreements than in the absence of these expectations.

The remainder of the paper is organized as follows. Section 2 briefly reintroduces the model. Section 3 studies the static Nash equilibrium of the model and then the dynamic game. Here, we first restate some of HIT's propositions as results, then refine them. Finally, Section 4 concludes. See HIT for the existence of the Nash equilibrium and the necessary technical assumptions, the equilibrium selection criteria, detailed interpretations of the assumptions used, and the proofs of their restated results. All the new proofs are relegated to Appendix.

## II. The Model

The world consists of two countries,  $A$  and  $B$ , and two goods,  $a$  and  $b$ . Country  $j$  is endowed with 1 unit of good  $-j$  and zero units of good  $j$ , where  $J \in \{A, B\}$  and  $J \in \{a, b\}$ . Demand functions are symmetric across countries and goods.  $D(P_j^J)$  is the demand for good  $j$  in country  $J$ , which depends only on good  $j$ 's price in country  $J$ ,  $P_j^J$ , satisfying the following standard assumptions: (i)  $D(P_j^J) > 0$  in  $[0, \bar{P}_j^J)$  and  $D(P_j^J) = 0$  for  $P_j^J \geq \bar{P}_j^J$ , (ii)  $D(P_j^J)$  is twice continuously differentiable in  $P_j^J$  with  $D'(P_j^J) < 0$  for  $P_j^J \in [0, \bar{P}_j^J)$ . Country  $J$  imports good  $j$  and exports good  $-j$  in accordance with the following export supply function:  $X_{-j}^J(P_{-j}^J) = 1 - D(P_{-j}^J)$ .

The countries have perfect information regarding the past choices and in each period they simultaneously select their (nonprohibitive) import tariffs to maximize their individual welfares. The no-arbitrage condition for good  $j$  yields  $P_j^J = P_j^{-J} + \tau^J$ , where  $\tau^J > 0$  is country  $J$ 's import tariff. Finally, the market-clearing conditions,  $D(P_j^J(\tau^J)) = X_{-j}^{-J}(P_{-j}^{-J}(\tau^J))$ , provide the equilibrium prices.

The countries have preferences for fairness and reciprocity and country  $J$ 's welfare is given by:

$$RW^J(\tau^J, \tau^{-J}, \tau_f^{-J}) = RW^J(\tau^J, \tau^{-J}) + \gamma w^J(\tau^{-J}, \tau_f^{-J}) SW^{-J}(\tau^J, \tau^{-J}). \quad (1)$$

The first term is the self-interested (or standard) welfare function, i.e., the sum of consumer surplus, producer surplus, and tariff revenue:

When do fairness and reciprocity hinder multilateral tariff cooperation? 7

$$\begin{aligned}
 SW^J(\tau^J, \tau^{-J}) &= \int_{P_j^J(\tau^J)}^{\bar{P}_j^J} D(P)dP \\
 &+ \int_{P_j^J(\tau^{-J})}^{\bar{P}_j^J} D(P)dP + P_j^J(\tau^{-J}) + \tau^J X_j^{-J}(\tau^J).
 \end{aligned} \tag{2}$$

The second term is the fairness payoff for country  $J$ , where (i)  $\gamma > 0$  is a scaling factor; and (ii)  $w^J(\tau^{-J}, \tau_f^J)$  determines the (scaled) weight country  $J$  places on its trading partner's self-interested welfare, has the following form:

$$w^J(\tau^{-J}, \tau_f^J) \begin{cases} \in (0, 1] \text{ if } \tau^{-J} > \tau_f^J \\ = 0 \text{ if } \tau^{-J} = \tau_f^J, \\ \in [-1, 0) \text{ otherwise} \end{cases} \tag{3}$$

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

Incorporating fairness payoff yields countries to not only care about the outcomes but also their trading partner's intentions. In brief, equation (3) indicates that from country  $J$ 's perspective, country  $-J$  that should be rewarded for its fair tariff, i.e., any  $\tau^{-J}$  below  $\tau_f^J$  (positive reciprocity), whereas it should be punished for its unfair tariff, i.e., any  $\tau^{-J}$  in excess of  $\tau_f^J$  (negative reciprocity.)

### III. Games

#### 1. Static Game

In this section, we first characterize the static Nash equilibrium of the game reciprocal game  $\Gamma^R(RW, w, \vec{\tau}_f)$ , where  $\vec{\tau}_f \equiv (\tau_f^J, \tau_f^{-J})$  is the fair-tariff vector, and compare it with the one that would emerge in a game with self-interested countries  $\Gamma^S(SW)$ . We henceforth assume that the countries have a common fair-tariff perception, i.e.,  $\tau_f^J = \tau_f^{-J} \equiv \tau_f$ , for simplicity.<sup>4)</sup>

In this setup, we have the following observations: (i) the choice variables in  $\Gamma^R(RW, w, \vec{\tau}_f)$  are strategic complements, i.e.,  $\partial^2 RW^J / \partial \tau^J \partial \tau^{-J} \geq 0$ , (ii) the cross-partial derivative of country  $J$ 's welfare function with respect to its tariff and its critical fair-tariff perception is nonpositive, i.e.,  $\partial^2 RW^J / \partial \tau^J \partial \tau_f \leq 0$ , (iii)  $\partial^2 SW^J / \partial \tau^J \partial \tau^{-J} = 0$  due to goods' demand independence of each other's price, and  $\partial^2 SW^J / \partial \tau^J \partial \tau_f = 0$  since  $SW^J$  is not a function of  $\tau_f$ .

Let  $\vec{\tau}_{NR} \equiv (\tau_{NR}, \tau_{NR})$  and  $\vec{\tau}_{NS} \equiv (\tau_{NS}, \tau_{NS})$  refer to the Nash equilibria of  $\Gamma^R(RW, w, \vec{\tau}_f)$  and  $\Gamma^S(SW)$ , respectively. We also assume that overly restrictive import policies are considered unfair, i.e.,  $\tau_f \leq \tau_{NS}$ , which is a reasonable assumption given our focus on trade cooperation among countries.

#### Result 1 (Static Game Results).

1.  $\vec{\tau}_{NR} \equiv (\tau_{NR}, \tau_{NR})$ , is non-increasing in  $\tau_f$ .

---

4) HIT also discusses the impact of asymmetries in fair-tariff perceptions.

When do fairness and reciprocity hinder multilateral tariff cooperation? 9

2. For  $\tau_f \leq \tau_{NS}$ , (i)  $\tau_{NR} \geq \tau_{NS}$  and (ii) for any  $J$ ,  
 $RW^J(\vec{\tau}_{NR}, \tau_f) \leq SW^J(\vec{\tau}_{NS})$ , with equality only holding for  
 $\tau_f = \tau_{NS}$ .

The intuition behind the first point is, for a given  $\tau^{-J}$ , a higher  $\tau_f$  leads to a larger  $w^J$ . As a result, country  $J$ , when choosing its import policy, has incentive to reduce the terms-of-trade negative externality of its tariff on its trading partner, resulting in a more liberal Nash tariff equilibrium. The intuition is behind the second point is, at  $\tau_{NS}$ , reciprocal countries are in a negative-reciprocity state having incentive to punish each other (since  $\tau_f \leq \tau_{NS}$ ), implying that the Nash equilibrium tariff of  $\Gamma^R(RW, w, \vec{\tau}_f)$  must exceed the one of  $\Gamma^S(SW)$ .

## 2. Dynamic Game

In the dynamic game, the stage game analyzed above is repeated infinitely many times. The aim in this section is to evaluate the effect of fairness and reciprocity on the ability of countries sustaining cooperative import tariffs when they are limited to self-enforcing trade agreements.

Let  $\Gamma_\infty^R(RW, w, \vec{\tau}_f)$  and  $\Gamma_\infty^S(SW)$  be the infinitely-repeated game with reciprocal and self-interested countries, respectively, and  $\delta \in (0, 1)$  be the discount factor between periods. For both games, given the overall symmetry of our framework, we focus on symmetric cooperative subgame-perfect equilibria, which is assumed to be common in both games. Also, countries employ a Nash reversion strategy to enforce cooperation. That is, (i) along the equilibrium path, the countries set a common cooperative tariff  $\tau_C < \tau_{NS} \leq \tau_{NR}$  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. Notice that the games we aim to compare,  $\Gamma_{\infty}^R(RW, w, \vec{\tau}_f)$  and  $\Gamma_{\infty}^S(SW)$ , are identical except the fairness and reciprocity.

Let us begin the analysis with  $\Gamma_{\infty}^S(SW)$ . The self-interested country  $J$  has the static incentive to cheat, that is, onetime gain by deviating from cooperative tariff:

$$SW^J(BR_S^J(\tau_C), \tau_C) - SW^J(\vec{\tau}_C) \equiv SW_D^J - SW_C^J \equiv \Omega_S^J(\tau_C), \quad (4)$$

where  $BR_S^J(\tau_C)$  is country  $J$ 's best-response tariff to  $\tau_C$  and  $\vec{\tau}_C \equiv (\tau_C, \tau_C)$ . On the other hand, violating multilateral cooperation ensues a trade war. The discounted future welfare cost a defector faces equals:

$$\frac{\delta}{1-\delta}(SW^J(\vec{\tau}_C) - SW^J(\vec{\tau}_{NS})) \equiv \frac{\delta}{1-\delta}(SW_C^J - SW_N^J) \equiv \frac{\delta}{1-\delta}\omega_S^J(\tau_C) \quad (5)$$

where  $\omega_S^J$  is the per-period value of cooperation for country  $J$ . Therefore, the incentive-compatibility condition for a self-interested country  $J$  to sustain cooperation is:

$$\Omega_S^J(\tau_C) \leq \frac{\delta}{1-\delta}\omega_S^J(\tau_C). \quad (6)$$

From ([ICS]), a given cooperative tariff  $\tau_C$  can be supported as a subgame-perfect equilibrium of  $\Gamma_{\infty}^S(SW)$  if countries are sufficiently patient:

When do fairness and reciprocity hinder multilateral tariff cooperation? 11

$$\delta \geq \delta_{\tau_C}^S \equiv \frac{SW_D^J - SW_C^J}{SW_D^J - SW_N^J}. \quad (7)$$

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

$$\Omega_R^J(\tau_C) \leq \frac{\delta}{1-\delta} \omega_R^J(\tau_C). \quad (8)$$

Moreover, for a given cooperative tariff  $\tau_C$ , the minimum discount factor needed to sustain cooperation at the agreed upon cooperative tariff is:

$$\delta_{\tau_C}^R \equiv \frac{RW_D^J - RW_C^J}{RW_D^J - RW_N^J}. \quad (9)$$

Using (7) and (9), we analyze the impact of fairness and reciprocity on multilateral cooperation by comparing  $\delta_{\tau_C}^S$  against  $\delta_{\tau_C}^R$ .

**Result 2** (HIT Proposition 2). *Suppose  $\gamma$  is sufficiently small.*

1. *For any  $\tau_C \leq \tau_f$ , we have  $\delta_{\tau_C}^R < \delta_{\tau_C}^S$ .*
2. *However, for any  $\tau_C > \tau_f$ , either  $\delta_{\tau_C}^R < \delta_{\tau_C}^S$  or  $\delta_{\tau_C}^R \geq \delta_{\tau_C}^S$  could hold.*

The assumption  $\gamma$  being sufficiently small means 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. HIT, in footnote 11, reports the difficulty in finding a closed-form solution for the upper bound of  $\gamma$ .<sup>5)</sup>

The intuition behind Result 2 is as follows. For any cooperative tariff  $\tau_C$  lower than the critical fair-tariff perception  $\tau_f$ , the countries are in a positive-reciprocity state. In such a case, two reinforcing forces are at work. First, for any country  $J$ , the value of cooperation at  $\tau_C$  is higher reciprocal game than the one in self-interested game,  $\omega_R^J > \omega_S^J$ , since in the former game (i) the noncooperative Nash tariff is higher; and (ii) while the cooperative tariff is common in both games, in the reciprocal game, country  $J$  would additionally enjoy a positive fairness payoff. Second, the static incentive country  $J$  has to deviate from  $\tau_C$  is weaker in reciprocal game than the one with self-interested game,  $\Omega_R^J < \Omega_S^J$ , because in the former game since the countries are in a positive-reciprocity state (i) countries enjoy a positive fairness payoff when they cooperate; (ii) the defect tariff is lower; and (iii) defection would hurt  $J$ 's partner, mitigating  $J$ 's potential onetime gains from cheating. Therefore, fairness and reciprocity facilitates the sustainability of a cooperative tariff lower than their perception of a critical fair-tariff.

However, for any cooperative tariff  $\tau_C$  higher than the critical fair-tariff perception  $\tau_f$ , the countries are in a negative-reciprocity state. For any such  $\tau_C > \tau_f$ , the static incentive country  $J$  has to

---

5) Rabin (1993). argues that the impact of reciprocal concerns should be much smaller as the size of the stakes increases. Given the large economic gains with multilateral trade agreements, assuming small  $\gamma$  is reasonable. However, to the best of my knowledge, there is no empirical evidence for this assumption in trade literature.

deviate from  $\tau_C$  becomes stronger in reciprocal game than the one with self-interested game,  $\Omega_R^J > \Omega_S^J$ , 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  $J$ . However, (iii) as they are willing to hurt country  $-J$ , country  $J$  would sacrifice some of its standard welfare, which would decrease their incentive to cheat. The two previous effects dominate the latter effect (iii) for sufficiently small  $\gamma$ . Finally, (iv) countries are also in a negative reciprocity state at  $\tau_C$ , which would decrease their welfare at cooperation, increasing their incentive to cheat. Since the incentive to cheat appears in the numerators of (7) and (9), reciprocal countries having stronger incentive to cheat yields a higher critical discount factor to sustain cooperation in  $\Gamma_\infty^R(RW, w, \vec{\tau}_f)$  than in  $\Gamma_\infty^S(SW)$ . On the other hand, since the reciprocal countries have higher welfare when they defect from cooperation as explained above and still lower welfare at the punitive Nash tariffs, the denominator of (9) would be higher than the denominator of (7), which yields a lower critical discount factor to sustain cooperation in  $\Gamma_\infty^R(RW, w, \vec{\tau}_f)$  than in  $\Gamma_\infty^S(SW)$ . Therefore, the impact of fairness and reciprocity on sustaining multilateral tariff cooperation becomes ambiguous.<sup>6)</sup>

They, in Proposition 3, obtain that similar finding regarding the most cooperative symmetric equilibria that can be supported by

---

6) In explaining this result, HIT provides similar intuition provided here and stated that the value of cooperation at  $\tau_C$  is higher in  $\Gamma_\infty^R(RW, w, \vec{\tau}_f)$  than in  $\Gamma_\infty^S(SW)$ . However, the denominators of (9) and (7) do not represent the value of cooperation at  $\tau_C$ . In fact, in Proposition 1, we are going to provide a sufficient condition for the the value of cooperation at  $\tau_C$  to be lower in  $\Gamma_\infty^R(RW, w, \vec{\tau}_f)$  than in  $\Gamma_\infty^S(SW)$ .

reciprocal and self-interested countries. Our first new result in this paper would generalize the Result 2 (and HIT's Proposition 3.)

**Proposition 1.** *The assumption, sufficiently small  $\gamma$ , is not necessary for Result 2.*

Although it is reasonable for  $\gamma$  to be small, Proposition 1 states that the Result 2 (and HIT's Proposition 3) will continue to hold for higher levels of fairness and reciprocity concerns. Since the weight function (3) can give values in  $[-1, 1]$ , the relative weight of the fairness payoff in the countries' objective function can be as high as their standard welfare.

Next, we provide sufficiency conditions for the impact of fairness and reciprocity to be detrimental on multilateral tariff cooperation.

**Proposition 2.** (Sufficiency Conditions). *Given a common cooperative tariff for reciprocal and self-interested games  $\tau_C > \tau_f$ , if (i) the scaling factor  $\gamma$  is sufficiently small and (ii) for any country  $J$  the fairness payoff at cooperative tariff is at least as small as the one at Nash tariffs, i.e.,  $w^J(\tau_C, \tau_f)SW_C^{-J} \leq w^J(\tau_{NR}, \tau_f)SW_N^{-J}$ , then we have  $\delta_{\tau_C}^R > \delta_{\tau_C}^S$ .*

The underlying intuition of the first part of Proposition 2 is as follows. First, recall that for any cooperative tariff  $\tau_C$  higher than the fair tariff  $\tau_f$ , the countries are in a negative-reciprocity state. Also, the static incentive country  $J$  has to deviate from  $\tau_C$  becomes stronger in reciprocal game than the one with self-interested game,  $\Omega_R^J > \Omega_S^J$ . Moreover, if (i) and (ii) holds, then, for any

country  $J$ , the value of cooperation at  $\tau_C$  becomes at least as high as in self-interested game than in reciprocal game,  $\omega_R^J \leq \omega_S^J$ , because the condition (ii) means that the impact of negative fairness payoffs will be at least as strong as at cooperative tariffs than at Nash tariffs. To gain more intuition behind this condition, notice that  $SW_C^J > SW_N^J$  and  $w^J(\tau_{NS}, \tau_f) \leq w^J(\tau_C, \tau_f) < 0$ . Thus, a sufficient condition for the condition (ii) to hold is the difference between the negative weight country  $J$  places on trading partner  $-J$ 's standard welfare at cooperative tariffs to be close to the one at non-cooperative Nash tariff, i.e.,  $|w^J(\tau_C, \tau_f) - w^J(\tau_{NR}, \tau_f)|$  to be sufficiently small. In other words, a reciprocal country  $J$  would consider trading partner country  $-J$ 's cooperative tariff almost as unkind as its Nash tariff. Therefore, both  $\Omega_R^J > \Omega_S^J$  and  $\omega_R^J \leq \omega_S^J$  hold and, thus, the impact of fairness and reciprocity would be detrimental on multilateral tariff cooperation under these sufficiency conditions. The same sufficiency conditions would modify HIT's Proposition 3 in similar manner.

## IV. Conclusions

This paper tunes the findings of HIT, which explores the impact of fairness and reciprocity on multilateral tariff cooperation. They find that reciprocal countries that are moderately demanding from their trading partners regarding their commercial policy (i.e., assuming the critical fair-tariff perception is not too low) can support a greater degree of multilateral tariff cooperation than self-interested ones. However, if reciprocal 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), the overall effect of fairness and reciprocity on multilateral tariff cooperation could be negative.

This paper first shows that these results hold even if the countries care for preferences for fairness and reciprocity as much as they care for their standard welfares. Second, it provides meaningful sufficient conditions for the fairness and reciprocity to hinder multilateral trade agreements when the countries are highly demanding from their trading partners. These conditions are: countries' fairness and reciprocity concerns are relatively small and their negative fairness payoff at a cooperative tariff does not exceed the one at Nash tariff. The latter condition is also satisfied if the trading partner's cooperative tariff is perceived almost as unkind as their non-cooperative Nash tariffs.

This paper and also HIT lack in several points. First, we have difficulty in finding empirical evidence of fairness and reciprocal concerns matter for the international trade negotiations. In HIT, we argue that fairness and reciprocal concerns are widespread among individuals and if the median voter has such concerns, then governments and country delegates should also care about them to be re-elected. We also provide many anecdotal evidence regarding the importance of such concerns. An interesting future work can be to conduct a survey study, similar to Dannenberg et al. (2017). They surveyed experts with different degrees of involvement in the climate negotiations and find that the ones that are more involved in the process are more optimistic about the assessment of the Paris Agreement.

Another interesting point that we did not address is the possibility of countries having disagreement in what they perceive kind and unkind. This would yield asymmetric Nash tariffs: the more demanding country having higher Nash tariffs than the less

demanding country. So, countries would have a different degree of punishment in the dynamic game. Assuming countries to still agree on a common cooperative tariff, which is reasonable given that they have identical material payoffs, the minimum discount factor needed to sustain cooperation for the one with the harsher punishment is lower than the one with the softer punishment. Therefore, for them to sustain the common cooperative tariff they have to be as patient as the less demanding country. While, I do not provide a detailed analysis here, in another working paper, I study this problem in a linear public goods game and a coalition formation game (İriş, 2024).

## Appendix

**Proof of Proposition 1:** In the proof of Result 2, HIT shows that

$$\tau_C \leq \tau_f \text{ implies that } \delta_{\tau_C}^R = ((RW_D^J - RW_C^J)/(RW_D^J - RW_N^J)) < ((SW_D^J - SW_C^J)/(SW_D^J - SW_N^J)) = \delta_{\tau_C}^S. \text{ To do so, HIT proves:}$$

1. If  $\tau_C \leq \tau_f \Rightarrow RW_D^J - RW_C^J \leq SW_D^J - SW_C^J$  for any  $J$ .
2. If  $\tau_C \leq \tau_f \Rightarrow RW_D^J - RW_N^J \leq SW_D^J - SW_N^J$  for any  $J$ .

The only part that HIT uses sufficiently small  $\gamma$  assumption is to prove,  $RW_D^J - SW_D^J \geq 0$  for any  $J$  in inequality (A3) on page 939. We are going to prove it without using the sufficiently small  $\gamma$  assumption.

$$\begin{aligned} RW^J(BR_R^J(\tau_C), \tau_C, \tau_f) &= \max_x [SW^J(x, \tau_C) + \gamma w^J(\tau_C, \tau_f) SW^{-J}(x, \tau_C)] \\ &\geq SW^J(BR_S^J(\tau_C), \tau_C) + \gamma w^J(\tau_C, \tau_f) SW^{-J}(BR_S^J(\tau_C), \tau_C) \\ &\geq SW^J(BR_S^J(\tau_C), \tau_C) \end{aligned} \tag{10}$$

The last equality is because  $w^J(\tau_C, \tau_f) \geq 0$  for  $\tau_C \leq \tau_f$ . *Q.E.D.*

**Proof of Proposition 1:** Let us begin the proof with the static incentive to cheat. Recall that now we assume  $\gamma$  to be sufficiently small. Taking a first-order Taylor series expansion of  $RW^J(BR_R^J(\tau_C), \tau_C, \tau_f) - RW^J(\vec{\tau}_C, \tau_f)$  around  $\gamma = 0$  and rearranging terms, we obtain:

$$\begin{aligned} RW^J(BR_R^J(\tau_C), \tau_C, \tau_f) - RW^J(\vec{\tau}_C, \tau_f) &- SW^J(BR_S^J(\tau_C), \tau_C) - SW^J(\vec{\tau}_C) \\ &\approx \gamma w^J(\gamma_C, \tau_f) (SW^{-J}(BR_S^J(\tau_C), \tau_C) - SW^{-J}(\vec{\tau}_C)) > 0. \end{aligned} \tag{11}$$

Because  $w^J(\tau_C, \tau_f) < 0$  for  $\tau_C > \tau_f$  and  $SW^{-J}(BR_S^J(\tau_C), \tau_C) - SW^{-J}(\vec{\tau}_C) < 0$  as the standard welfare of the trading partner  $-J$  when  $J$  defects is lower than the one when  $J$  continues to cooperate. Thus, if  $\gamma$  is sufficiently small and  $\tau_C > \tau_f$ , we have  $\Omega_R^J > \Omega_S^J$ .

Let us continue with the value of cooperation. Taking a first-order Taylor series expansion of  $RW^J(\vec{\tau}_C, \tau_f) - RW^J(\vec{\tau}_{NR}, \tau_f)$  around  $\gamma = 0$  and rearranging terms, we obtain:

$$\begin{aligned} RW^J(\vec{\tau}_C, \tau_f) - RW^J(\vec{\tau}_{NR}, \tau_f) - SW^J(\vec{\tau}_C) + SW^J(\vec{\tau}_{NS}) & \quad (12) \\ \approx \gamma(w^J(\tau_C, \tau_f)SW^{-J}(\vec{\tau}_C) - w^J(\tau_{NS}, \tau_f)SW^{-J}(\vec{\tau}_{NS})). \end{aligned}$$

First notice that  $SW_C^{-J} > SW_N^{-J} > 0$ . Moreover, for  $\tau_{NS} > \tau_C > \tau_f$ , we have  $w^J(\tau_{NS}, \tau_f) \leq w^J(\tau_C, \tau_f) < 0$ . Thus, if the difference between the weights at cooperative and Nash tariffs, i.e.,  $|w^J(\tau_C, \tau_f) - w^J(\tau_{NR}, \tau_f)|$ , is also sufficiently small, then the term in the second line of (12) would become negative for sure. Therefore, under these sufficiency conditions the value of cooperation at  $\tau_C$  would be lower in reciprocal game than in self-interested game, i.e.,  $\omega_R^J < \omega_S^J$ . The proof of the second part of Proposition 1 is immediate from the proof of Result 3. *Q.E.D.*

## References

1. Bagwell, Kyle, and Robert Staiger (1999) "Regionalism and multilateral tariff cooperation," in *International Trade Policy and the Pacific Rim*, ed. J. Piggott and A. Woodland (London: Macmillan)
2. Camerer, Colin (2003) "Behavioral Game Theory. Experiments in strategic interaction," Princeton University Press/Russell Sage Foundation.
3. Dannenberg, Astrid, Sonja Zitzelsberger, and, Alessandro Tavoni (2017) "Climate negotiators' and scientists' assessments of the climate negotiations," *Nature Clim Change* 7, 437-442.
4. Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde (2009) "Homo reciprocans: Survey evidence on behavioural outcomes," *Economic Journal* 119, 592-612
5. Falk, Armin, Ernst Fehr, and Urs Fischbacher (2003) "On the nature of fair behavior," *Economic Inquiry* 41(1), 20-26.
6. Falk, Armin, Ernst Fehr, and Urs Fischbacher (2008) "Testing theories of fairness â intentions matter," *Games and Economic Behavior* 62(1), 287-303.
7. Fehr, Ernst, and Simon Gächter (2002) "Altruistic punishment in humans," *Nature* 415, 137-140.
8. Fehr, Ernst, Simon Gächter, and Georg Kirchsteiger (1997) "Reciprocity as a contract enforcement device: Experimental evidence," *Econometrica* 65, 833-860
9. Fehr, Ernst, Elena Tougareva, and Urs Fischbacher (2014) "Do high stakes and competition undermine fair behaviour? Evidence from Russia," *Journal of Economic Behavior & Organization*, 108, 354-363.
10. Hadjiyiannis, Costas, Doruk İriş, and Chrysostomos Tabakis (2012a) "Multilateral tariff cooperation under fairness and reciprocity," *Canadian Journal of Economics*, 45(3), 925-941.
11. Hadjiyiannis, Costas, Doruk İriş, and Chrysostomos Tabakis (2012b) "International environmental cooperation under fairness and reciprocity," *B.E. Journal of Economic Analysis and Policy*,

- 12(1).
12. İriş, Doruk (2015) "Cournot duopoly and tacit collusion under fairness and reciprocal preferences," *Journal of Economic Theory and Econometrics*, 26(3), 1-18.
  13. İriş, Doruk (2024) "Expectations of Reciprocal Negotiators," *Mimeo*.
  14. İriş, Doruk (2018) "Tacit collusion under fair price perceptions and reciprocal preferences: The role of expectations from rivals," *Korean Review of Applied Economics*, 20(3), 145-182.
  15. İriş, Doruk, and Suha Kim (2024) "Reciprocal Preferences and Expectations in International Agreements," *SSRN Working Paper*, <https://ssrn.com/abstract=4788495>.
  16. İriş, Doruk, and Lu í's Santos-Pinto (2013) "Tacit collusion under fairness and reciprocity," *Games*, 4(1), 50-65.
  17. İriş, Doruk, and Alessandro Tavoni (2020) "Tipping and Reference Points in Climate Change Games," *Handbook on the Economics of Climate Change*, eds. Chichilnisky and Rezai, Edward Elgar Press, UK.
  18. Kocher, Martin G., Peter Martinsson, and Martine Visser (2008) "Does stake size matter for cooperation and punishment?," *Economics Letters*, 99(3), 508-511.
  19. Lange, Andreas (2006) "The impact of equity-preferences on the stability of international environmental agreements," *Environmental and Resource Economics* 34, 247-267.
  20. Lange, Andreas, Carsten Vogt (2003) "Cooperation in international environmental negotiations due to a preference for equity," *Journal of Public Economics* 87, 2049-2067.
  21. Larney, Andrea, Amanda Rotella, and Pat Barclay (2019) "Stake size effects in ultimatum game and dictator game offers: A meta-analysis," *Organizational Behavior and Human Decision Processes*, 151, 61-72.
  22. Lü, Xiaobo, Kenneth Scheve, and Matthew J. Slaughter (2012) "Inequity aversion and the international distribution of trade protection," *American Journal of Political Science* 56(3), 638-654.

22 시장경제연구 53호 3집

23. Nyborg, Karine (2017) "Reciprocal climate negotiators," *Journal of Environmental Economics and Management*, 92, 707-725.
24. Rabin, Matthew (1993) "Incorporating fairness into game theory and economics," *American Economic Review* 83, 1281-1302
25. Sobel, Joel (2005) "Interdependent preferences and reciprocity," *Journal of Economic Literature* 43, 392-436.
26. Segal, Uzi, and Joel Sobel (2007) "Tit for tat: Foundations of preferences for reciprocity in strategic settings," *Journal of Economic Theory* 136, 197-216

국문초록

## 공정성과 상호주의가 다자간 관세 협력을 저해하는 시기는 언제인가?

Doruk İriş\*

이 논문은 공정성과 상호 호혜가 다자간 관세 협력에 미치는 영향을 연구한다. 상호 호혜적인 국가는 친절한 행동에 보답하는 반면(긍정적 상호 호혜), 불친절한 행동을 보인 국가를 처벌한다(부정적 상호 호혜). 이전 연구에 따르면, 공정성과 상호 호혜의 영향이 상대적으로 작을 때, 무역 파트너의 상업 정책에 대해 적당한 요구를 할 경우 상호 호혜는 다자간 관세 협력을 촉진한다. 반면, 매우 자유로운 수입 정책만을 공정하다고 여길 때는 협력에 부정적인 영향을 미칠 수 있다. 그러나 본 논문은 첫째로, 공정성과 상호 호혜의 영향이 크더라도 이러한 결과가 여전히 유효하다는 것을 보인다. 둘째로, 공정성과 상호 호혜가 다자간 관세 협력을 저해할 수 있는 충분한 조건을 제공한다.

**핵심주제어:** 호혜성, 공정성, 무역 계약, 반복 게임

**JEL Classification:** F13; D90; D63.

---

\* 서강대학교 경제학부, Email: dorukiris@gmail.com.