

Journal of International Economics 61 (2003) 397-424

Journal of INTERNATIONAL ECONOMICS

www.elsevier.com/locate/econbase

# Discriminatory tariffs and international negotiations

Josh Ederington<sup>a,\*</sup>, Phillip McCalman<sup>b</sup>

<sup>a</sup>Department of Economics, University of Miami, P.O. Box 248126, Coral Gables, FL 33124, USA <sup>b</sup>University of California, Santa Cruz, CA, USA

Received 25 June 2002; received in revised form 11 July 2002; accepted 16 September 2002

#### Abstract

Recent research has highlighted the efficiency of the MFN principle within the GATT/ WTO structure. This paper analyzes the exception made to MFN within Article XXIII that allows discriminatory punishment for deviations from the agreement. We argue that, in the absence of collusion, the MFN exception reduces the severity of punishment and thus lowers the level of cooperation that can be achieved by the agreement. However, discriminatory punishment may still be beneficial as we show that it reduces the problems associated with the potential for renegotiation during the punishment phase. Finally, we argue that our results are also applicable to the question of whether to use trade policy sanctions as a means of enforcing agreements covering domestic policies. © 2003 Elsevier B.V. All rights reserved.

Keywords: Trade agreements; Dispute settlement; Non-discrimination

JEL classification: F02; F13; F15

# 1. Introduction

Since its inception, the General Agreement on Tariffs and Trade (GATT) has been extremely successful in liberalizing trade barriers. One of the pillars of the GATT system is the most-favored-nation (MFN) principle that requires each member of GATT to offer market access on non-discriminatory terms to all other

<sup>\*</sup>Corresponding author. Tel.: +1-305-284-1626; fax: +1-305-284-2985. E-mail address: ederington@miami.edu (J. Ederington).

<sup>0022-1996/03/\$ -</sup> see front matter © 2003 Elsevier B.V. All rights reserved. doi:10.1016/S0022-1996(03)00004-7

members of GATT. Recent papers (e.g., Bagwell and Staiger, 1999a, 2002; McCalman, 2002; Zissimos and Vines, 1999) have made clear that this principle of non-discrimination plays a key role in the achievement of efficient multilateral trade agreements. Given these efficiency properties, it seems beneficial to analyze the cases under which exceptions to MFN are allowed in GATT.

There are three main cases where exceptions to MFN are made within the GATT framework. The first involves 'safeguard' actions under Article XIX (the escape clause), although departures from MFN under Article XIX are only allowed on consultation with the Safeguards Committee. The second involves the signing of preferential customs unions or free trade agreements under Article XXIV. The third main exception involves retaliatory actions under the dispute settlement mechanism of Article XXIII. If GATT authorizes a retaliatory action under Article XXIII, such action need not be taken on MFN basis (see Jackson, 1989). It is this last exception that is the subject of this paper.

The exception to MFN under Article XXIII allows countries to discriminate in the dispute settlement phase of the agreement. Thus, under Article XXIII, in the event that a country deviates from the agreement, member countries are allowed to punish only the deviating country while maintaining cooperation among themselves. The question raised by this paper is whether allowing such discriminatory punishment can assist in enforcing greater cooperation within the agreement.

We adopt the view that enforcement issues are central to the design and understanding of international agreements. Specifically we argue that, in the absence of an external enforcement mechanism, international agreements are only viable if member countries view continued cooperation to be in their own self-interest. The GATT/WTO dispute settlement mechanism may play an important role in coordinating multilateral enforcement efforts (see, for example, Maggi, 1999). However, it has no independent ability to punish violations on its own. Thus, international agreements like GATT are limited in their ability to enforce cooperation by the severity of punishment which member countries can credibly threaten against potential cheaters. In Section 3 of the paper we argue that discriminatory punishment actually reduces the severity of punishment that can be threatened to potential cheaters, and thus leads to a less cooperative agreement.<sup>1</sup> Thus, countries would be better off if they could credibly commit to nondiscriminatory punishment (symmetric punishment of all members of the agreement) since such punishment is a stronger deterrent to deviations. Such a result casts doubts on the efficiency of the Article XXIII exception to MFN.

However, in Sections 4 and 5 of the paper we provide two arguments that

<sup>&</sup>lt;sup>1</sup>This result is driven by a 'tariff complementarity' effect whereby a reduction in the tariff against one country results in a reduction in the optimal tariff against another country. Thus, our result is comparable to the recent theoretical literature on how the formation of trading blocs affects the ability to sustain multilateral agreements where tariff complementarity also plays an important role (e.g., Bagwell and Staiger, 1999b; Bond et al., 2001).

suggest discriminatory punishment may be optimal within a self-enforcing agreement. In Section 4 we argue that discriminatory punishment is optimal when member countries are allowed to collude in setting tariffs against the cheating country during the punishment phase. In Section 5, we argue that, even in the absence of collusion, discriminatory punishments may be optimal when the self-enforcing agreement is susceptible to renegotiation. Indeed, as mentioned by Ludema (2001), the GATT dispute settlement mechanism actually encourages renegotiation after disputes have been triggered. As has been well established in the game theory literature, requiring punishments to be renegotiation-proof can reduce the severity of the punishment that is threatened to potential deviators, creating problems for the enforcement of agreements. In Section 5, we show, using a standard definition of renegotiation-proof punishments suggested by Ludema (2001), that allowing discriminatory punishments can reduce the problems associated with the potential for renegotiation during the punishment phase of the agreement. Thus, we argue that the Article XXIII exception to MFN is justified in agreements where punishments are sensitive to renegotiation.

Finally, in Section 6, we argue that our results are also relevant to the question of whether to use trade policy as a means of enforcing international agreements covering domestic policies (e.g., environmental policy). Specifically, we argue that one of the inherent advantages of trade policy over domestic policy as an instrument of enforcement is its discriminatory potential (i.e., trade sanctions can be used to punish only those members who cheat on the agreement, maintaining cooperation among the rest). The results of our paper suggest that discriminatory trade policy sanctions can be beneficial as a means of enforcing international domestic policy agreements that are sensitive to the problems of renegotiation.

# 2. The model

The analysis is conducted within a three-country (*X*, *Y* and *Z*), three-good (*x*, *y* and *z*) 'competing exporters' model of trade analagous to that employed by Bagwell and Staiger (1999b). Demand for good *i* in country *J* is given by a demand function  $D_J^i(p_J^i)$  where  $p_J^i$  is the local price of the good and  $D'(p_J^i) < 0$ . Production of good *i* in country *J* is defined by a supply function  $Q_J^i(p_J^i)$  where  $Q'(p_J^i) > 0$ .<sup>2</sup> These demand and supply functions are assumed to be defined so that each country is the 'natural' exporter of two of the goods and the unique importer of the third. For notational simplicity we let the lower case letter denote the unique import good for each country (e.g., country *Y* imports *y* and exports *x* and *z*).

<sup>&</sup>lt;sup>2</sup>Note that demand and supply are solely functions of own price effects and thus we are abstracting from cross-price and income effects. This framework can be rationalized in general equilibrium terms by assuming the presence of a numeraire good with quasi-linear utility and with production assumed to be a function of labor where labor supply is infinitely elastic at a unitary wage.

#### 400 J. Ederington, P. McCalman / Journal of International Economics 61 (2003) 397-424

We assume that countries choose specific import tariffs on their import goods which create a wedge between local prices in the importing and exporting countries. For example, define  $p_y^y$  as the local price (paid by consumers) of good y in the importing country Y. In that case, the local price of y in countries X and Z (the exporters of y) will be given (provided trade taxes are not prohibitive) by  $p_X^y(p_Y^y, \tau_{YX}) = p_Y^y - \tau_{YX}$  and  $p_Z^y(p_Y^y, \tau_{YZ}) = p_Y^y - \tau_{YZ}$  where  $\tau_{ii}$  represents the tariff placed by country i on the exports of country j. Prices of the other two goods are defined accordingly.<sup>3</sup>

Define excess demand for good *i* in country J by  $M_J^i(p_J^i)$  where  $M'(p_J^i) < 0$ . From the market clearing condition (i.e.,  $\Sigma_I M_I^i = 0$ ) and the demand and supply equations, one can derive the local price of each good as a function of tariff policy. Implicitly differentiating the market clearing condition, one can derive that an increase in tariffs against either of the two exporting countries will raise the local price of the good in the importing country. For example, with respect to good y:

$$\frac{\partial p_Y^{\nu}}{\partial \tau_{YX}} = \frac{M_X^{\nu'}}{\sum_J M_J^{\nu'}} > 0 \quad \text{and} \quad \frac{\partial p_Y^{\nu}}{\partial \tau_{YZ}} = \frac{M_Z^{\nu'}}{\sum_J M_J^{\nu'}} > 0 \tag{1}$$

From (1) and the definition of local prices, one can derive that an increase in the tariff against one exporting country will decrease the local price in that country, but will increase the local price in the alternate export country (i.e., the model exhibits trade diversion). For example, with respect to good y:

$$\frac{\partial p_X^y}{\partial \tau_{YX}} = \frac{-\left(M_Y^{y'} + M_Z^{y'}\right)}{\sum_J M_J^{y'}} < 0 \quad \text{and} \quad \frac{\partial p_Z^y}{\partial \tau_{YX}} = \frac{M_X^{y'}}{\sum_J M_J^{y'}} > 0 \tag{2}$$

Finally, governments are assumed to maximize the sum of consumer surplus and producer surplus, and tax revenue. Thus welfare for country Y on its import and export goods, respectively, is given by:<sup>4</sup>

$$W_{Y}^{y}(\tau_{YX}, \tau_{YZ}) \equiv \int_{\hat{\rho}_{Y}^{y}}^{\infty} D(\hat{p}) d\hat{p} + \int_{0}^{\hat{\rho}_{Y}^{y}} Q(\hat{p}) d\hat{p} - \tau_{YX} M_{X}^{y} - \tau_{YZ} M_{Z}^{y}$$

$$W_{Y}^{x}(\tau_{XY}, \tau_{XZ}) \equiv \int_{\hat{\rho}_{Y}^{x}}^{\infty} D(\hat{p}) d\hat{p} + \int_{0}^{\hat{\rho}_{Y}^{x}} Q(\hat{p}) d\hat{p}$$

$$W_{Y}^{z}(\tau_{ZY}, \tau_{ZX}) \equiv \int_{\hat{\rho}_{Y}^{y}}^{\infty} D(\hat{p}) d\hat{p} + \int_{0}^{\hat{\rho}_{Y}^{y}} Q(\hat{p}) d\hat{p}$$
(3)

<sup>3</sup>Given  $p_X^x$ ,  $p_Y^x(p_X^x, \tau_{XY}) = p_X^x - \tau_{XY}$  and  $p_Z^x(p_X^x, \tau_{XZ}) = p_X^x - \tau_{XZ}$ . Likewise, given  $p_Z^z$ ,  $p_Z^z(p_Z^z, \tau_{ZX}) = p_X^x - \tau_{XZ}$ .  $p_Z^z - \tau_{ZX}$  and  $p_Y^z(p_Z^z, \tau_{ZY}) = p_Z^z - \tau_{ZY}$ . <sup>4</sup>Welfare for countries X and Z are defined symmetrically.

In the absence of an international agreement, each country sets trade taxes on its importing good to maximize national welfare, taking the policy choices of its trading partner as given.

Taking the derivatives of (3) with respect to  $\tau_{YX}$  and  $\tau_{YZ}$ , and solving out the first-order conditions yields:

$$\tau_{YX}^{D} = \epsilon_{X}^{y} + (\tau_{YZ} - \epsilon_{Z}^{y}) \frac{M_{Z}^{y'}}{M_{Z}^{y'} + M_{Y}^{y'}}$$

$$\tau_{YZ}^{D} = \epsilon_{Z}^{y} + (\tau_{YX} - \epsilon_{X}^{y}) \frac{M_{X}^{y'}}{M_{X}^{y'} + M_{Y}^{y'}}$$
(4)

where  $\epsilon_i^i = M_J^i / M_J^{i'}$  is inversely related to the trade elasticity of good *i* from country J. The above two first-order conditions specify unilaterally optimal tariffs by country Y on imports from each country ( $\tau_{YX}^{D}$  and  $\tau_{YZ}^{D}$ ) and reflect familiar terms-of-trade incentives in which countries impose a positive tariff on imports so as to lower the world price of the good. Thus, from (4), the higher the trade elasticity of the exporting country, the lower the optimal tariff on imports from that country. However, when importing from multiple sources the terms of trade argument needs to be augmented. Note from (4) that the optimal tariff against each country is also a function of the tariff that is set against the alternate trading partner. This is due to the trade diversion exhibited in the model. For example, consider an increase by country Y of its tariff against country X ( $\tau_{YX}$ ). Due to trade diversion this leads to increased demand for the output of country Z and hence raises the local price of the good from country Z. To the extent that country Y is setting low tariffs against country Z (i.e.,  $\tau_{YZ} \le \epsilon_Z^y$ ) this increase in the price of the import good represents a terms of trade loss for country Y and thus reduces the incentive to raise tariffs against country X. Note that these terms-of-trade losses will be larger the more inelastic is trade with the alternate country (i.e., holding  $\tau_{yz}$ constant an increase in  $\epsilon_Z^y$  reduces the incentive to raise  $\tau_{yx}$ ). In the following Lemma we establish that, as a result of the above incentives, tariffs will be complementary. That is, an increase in the tariff against country X will lead to an increase in the optimal tariff against country Z:

**Lemma 1.** Assuming demand and supply functions are linear, a country's tariffs on imports of the same good from different trading partners will be complementary (i.e.,  $\partial \tau_{YX}^D / \partial \tau_{YZ} > 0$ ).

**Proof.** Taking the derivative of  $\tau_{YX}^{D}$  (defined by (4)) with respect to  $\tau_{YZ}$ , one derives that:<sup>5</sup>

<sup>&</sup>lt;sup>5</sup>Note that (5) is derived by either assuming that demand and supply functions are linear (so that second derivatives can be suppressed) or by evaluating the derivative in the area of the Nash equilibrium (where  $\tau_{yz} = \epsilon_z^{\gamma}$ ).

402 J. Ederington, P. McCalman / Journal of International Economics 61 (2003) 397-424

$$\left[1 - \frac{\partial \epsilon_X^y}{\partial \tau_{YX}} + \left(\frac{M_Z^{y'}}{M_Z^{y'} + M_Y^{y'}}\right) \frac{\partial \epsilon_Z^y}{\partial \tau_{YX}}\right] \frac{\partial \tau_{YX}}{\partial \tau_{YZ}} = \frac{\partial \epsilon_X^y}{\partial \tau_{YZ}} + \left(1 - \frac{\partial \epsilon_Z^y}{\partial \tau_{YZ}}\right) \frac{M_Z^{y'}}{M_Z^{y'} + M_Y^{y'}}$$
(5)

Simplifying using (1) and (2) and assuming that demand and supply functions are linear so that we can ignore second derivatives, one derives that:

$$\frac{\partial \tau_{YX}}{\partial \tau_{YZ}} = \frac{M_Z^{y'}(2\sum_j M_J^{y'} + M_Y^{y'})}{(M_Z^{y'} + M_Y^{y'}) \left[\sum_j M_J^{y'} + (M_Z^{y'} + M_Y^{y'}) + \frac{M_Z^{y'} \cdot M_X^{y'}}{(M_Z^{y'} + M_Y^{y'})}\right]}$$
(6)

It is direct to derive that, since both the numerator and denominator of the above expression are unambiguously positive, that tariffs are complementary.  $\Box$ 

The non-cooperative (Nash) equilibrium is the intersection of the above bestresponse functions and is given by:

$$\tau_{YX}^N = \epsilon_X^y \quad \text{and} \quad \tau_{YZ}^N = \epsilon_Z^y$$
(7)

Globally efficient trade policies will be set to maximize world welfare  $(W_X + W_Y + W_Z)$ , and will serve as the natural goals towards which countries strive when they cooperate. Taking the derivative of world welfare with respect to  $\tau_{YX}$  and  $\tau_{YZ}$  solving out the first-order conditions, we derive that globally optimal policies are defined by free trade ( $\tau_{YX} = \tau_{YZ} = 0$ ).

The setup of the model reflects the typical prisoners dilemma, in which countries want to cooperate to free trade, but have a unilateral incentive to erect barriers to trade. Unfortunately, the desire to erect trade barriers does not disappear once an agreement is in place, and a critical problem faced by any international agreement is the lack of an external enforcement mechanism to ensure that the signatories uphold their obligations. In the absence of external enforcement mechanisms, an agreement will only be viable if it is self-enforcing (i.e., member countries must view continued cooperation to be in their own best interest). Dixit (1987), Bagwell and Staiger (1990) and Riezman (1991) show how countries can support low tariffs within a repeated relationship by configuring the tariff agreement so that each country fears that a decision to cheat on the agreement would trigger a costly retaliatory episode in the future. In the following section, we employ this framework to determine if the discriminatory power of tariffs can facilitate cooperation.

#### 3. Self-enforcing agreements

An international agreement is defined by a set of binding cooperative policies (denoted  $\tau^c$ ). However, given the lack of an external enforcement mechanism, we require our agreement to be configured so that these cooperative trade barriers are self-enforcing. To that end, we assume that any deviation from these cooperative policies triggers a retaliatory episode. Thus, we characterize time periods as being either a period of 'cooperation', 'deviation' or 'retaliation'. The previous literature on self-enforcing international trade agreements (such as GATT) has focused on trigger strategies in which retaliation entails reversion, for a fixed number of periods, to the static Nash equilibrium (e.g., see Bagwell and Staiger, 1990; Riezman, 1991; Hungerford, 1991). The advantage of this approach is it ensures that the equilibrium of the repeated game is subgame perfect. Thus, as a benchmark case, we focus on credible (subgame perfect) agreements in which countries threaten to revert to the static Nash equilibrium in the event of a violation.

For tractability we will assume that our model is symmetric which allows us to conduct the analysis with respect to a single deviating country *X*, with symmetric conditions holding for the other two countries.<sup>6</sup> During periods of cooperation, countries set a low, common trade barrier (i.e.,  $\tau_{ij} = \tau^c$  for all countries). Given the symmetric nature of the model, common cooperative trade barriers imply that countries split the gains from cooperation equally. The level of welfare in a period of cooperation for country *i* is defined by:

$$W_i^c \equiv W_i^x(\tau_{XY} = \tau_{XZ} = \tau^c) + W_i^y(\tau_{YX} = \tau_{YZ} = \tau^c) + W_i^z(\tau_{ZX} = \tau_{ZY} + \tau^c).$$
(8)

However, provided that  $\tau^c < \tau^N$ , each country will have an incentive to deviate from this common cooperative policy. In this paper, we consider the case where a country deviates against both trading partners.<sup>7</sup> When deviating, a country will impose unilaterally optimal trade taxes (defined by (7)) and the level of welfare in a period of deviation (where country X has deviated on the agreement) for country i is given by:

$$W_i^D \equiv W_i^x(\tau_{XY} = \tau_{XZ} = \tau^N) + W_i^y(\tau_{YX} = \tau_{YZ} = \tau^c) + W_i^z(\tau_{ZX} = \tau_{ZY} = \tau^c).$$
(9)

Once a country deviates from the agreement, it triggers retaliation by the

<sup>&</sup>lt;sup>6</sup>Symmetry can be imposed by assuming that countries have identical demand and supply functions with respect to their respective import goods, left-hand side export goods and right-hand side export goods.

<sup>&</sup>lt;sup>7</sup>This can be justified by assuming that a violation (raising of a tariff against either of the two exporting countries) triggers multilateral retaliation by both countries (see Maggi, 1999 for a discussion of the issues relating to multilateral retaliation within international trade agreements). Given that any deviation by the importing country will trigger multilateral retaliation, a country will always cheat on both trading partners.

remaining countries. If this retaliation is non-discriminatory, then retaliation implies an abandonment of the entire agreement and a symmetric reversion to the Nash equilibrium on the part of all countries. Thus, with non-discriminatory punishment, countries will play Nash tariffs ( $\tau^N$  given by (7)) during a period of retaliation. Welfare for country *i* in a retaliatory period with nondiscriminatory punishment (denoted by NP) is given by:

$$W_i^{NP} \equiv W_i^x(\tau_{XY} = \tau_{XZ} = \tau^N) + W_i^y(\tau_{YX} = \tau_{YZ} = \tau^N) + W_i^z(\tau_{ZX} = \tau_{ZY} = \tau^N).$$
(10)

Alternatively, retaliation can be discriminatory, in which case non-cheating members of the agreement raise their tariffs against the cheating member while keeping tariffs low against each other. For example, assume country *X* has cheated on the agreement. Then, as a means of punishment, countries *Y* and *Z* will only revert to unilaterally optimal tariffs on trade with country *X*. That is, during a period of retaliation, taxes set on goods from country *X* are unilaterally optimal taxes for countries *Y* and *Z* (denoted  $\tau^D$  and defined by (4)), conditional on the fact that countries *Y* and *Z* are continuing to cooperate among themselves (i.e.,  $\tau_{YZ} = \tau_{ZY} = \tilde{\tau}^c$ )). We denote the lower trade barrier between countries *Y* and *Z* by  $\tilde{\tau}^c$  since there is no requirement that countries *Y* and *Z* must necessarily set the same cooperative tariffs,  $\tau^c$  as in the multilateral agreement. The only requirement we place is that  $\tilde{\tau}^c < \tau^{N.8}$ 

Therefore, provided that country X has deviated from the agreement in a past period, welfare for country i in a period of retaliation with discriminatory punishment (denoted by DP) is given by:

$$W_{i}^{DP} \equiv W_{i}^{x}(\tau_{XY} = \tau_{XZ} = \tau^{N}) + W_{i}^{y}(\tau_{YX} = \tau^{D}(\tilde{\tau}^{c}), \tau_{YZ} = \tilde{\tau}^{c}) + W_{i}^{z}(\tau_{ZX} = \tau^{D}(\tilde{\tau}^{c}), \tau_{ZY} = \tilde{\tau}^{c}).$$
(11)

An optimal international agreement results in the countries jointly choosing cooperative policies ( $\tau^c$ ) to maximize the cooperative level of welfare subject to the constraint that no country has an incentive to defect from the agreement. The  $\tau^c$  that satisfies this constrained maximization will be the smallest  $\tau^c$  that satisfies the self-enforcement constraint and will be referred to as the 'most-cooperative' tariff ( $\hat{\tau}^c$ ). Satisfying the self-enforcement constraint entails balancing the current gains to deviating from the agreement against the future longterm losses to retaliation. We assume that, in the case of a violation of the agreement, countries revert to retaliatory tariffs for T periods followed by a reversion to most-

<sup>&</sup>lt;sup>8</sup>Intuitively, we are assuming that the non-deviating countries maintain some degree of cooperation (i.e., low tariffs) among themselves. If  $\tilde{\tau}^c = \tau^N$ , then one would simply have the case of non-discriminatory punishment.

cooperative tariffs ( $\hat{\tau}^c$ ) thereafter, and refer to this sequence of periods as the 'punishment path'.

The self-enforcement constraint is satisfied provided that the discounted value of cooperating in the agreement  $(1/(1 - \delta)W^{C})$  is greater than the discounted value of deviating from the agreement and then triggering the punishment path. In the case of non-discriminatory punishment, the self-enforcement constraint for country *X* is given by:

$$g(\tau^{c})^{NP} \equiv \frac{1}{1-\delta} W_{X}^{C} - \left[ W_{X}^{D} + \frac{\delta}{1-\delta} ((1-\rho)W_{X}^{NP} + \rho W_{X}^{C}) \right] \ge 0$$
(12)

where  $\rho = \delta^T$  represents the relative weight placed on the cooperative versus non-cooperative welfare in the punishment path (and is, of course, a function of the duration of punishment, *T*).

In the case of discriminatory punishment, the self-enforcement constraint for country X is given by:

$$g(\tau^{c})^{DP} \equiv \frac{1}{1-\delta} W_{X}^{C} - \left[ W_{X}^{D} + \frac{\delta}{1-\delta} ((1-\rho)W_{X}^{DP} + \rho W_{X}^{C}) \right] \ge 0.$$
(13)

If the threat of retaliation is not sufficient to support free trade as a selfenforcing equilibrium (e.g., when countries discount the future heavily) then the self-enforcement constraint will bind. In that case, the question becomes which form of punishment (i.e., discriminatory or non-discriminatory) can support a more cooperative agreement.

One aspect of self-enforcing agreements is that the degree of cooperation that can be sustained is tied to the degree of punishment that can be threatened in the case of deviation. Specifically, more stringent punishment lowers the gain to deviating from the agreement and thus allows the agreement to support a more cooperative outcome. This is reflected in Lemma 2.

**Lemma 2.** Given that the self-enforcement constraint binds, a more stringent punishment (i.e., a lower level of welfare for the deviating country in a period of retaliation) can support a lower cooperative tariff.

# **Proof.** In Appendix A.

The above Lemma is reflected in Fig. 1. Rearranging (12) one derives that a cooperative tariff is supported within the non-discriminatory agreement provided that average welfare along the punishment path  $((1 - \rho)W_X^{NP} + \rho W_X^C)$  is less than  $\phi(\tau^c) \equiv W^C - [(1 - \delta)/\delta](W^D - W^C)$ . If this condition holds, then the gain to remaining in the agreement outweighs the potential gain to deviating on the agreement and then reverting to the punishment path. As is shown in Fig. 1, the range of enforceable tariffs, for which no country would have an incentive to deviate from the agreement, is given by the interval  $[\tau^{NP}, \tau^N]$ . Since global welfare



Fig. 1. Determining the most-cooperative tariff.

is monotonically increasing as the cooperative tariff is lowered towards free trade,  $\tau^{NP}$  represents the most-cooperative tariff  $(\hat{\tau}^c)$ .

Note from Fig. 1 that, holding  $\tau^c$  constant, as average welfare on the punishment path falls (i.e., the line denoted  $(1 - \rho)W_X^{NP} + \rho W_X^C$  shifts down) the agreement will be able to support lower cooperative tariffs (i.e.,  $\hat{\tau}^c$  will decrease). Lemma 2 implies that the optimal punishment scheme is the one that can threaten the most stringent punishment (i.e., the lowest level of welfare along the punishment path for the deviating country). In the analysis that follows, we determine which punishment path (discriminatory or non-discriminatory) is associated with the most stringent punishment for the deviating country.

Note, from (3), that foreign tariffs only affect home country welfare in their effect on export prices, and that home welfare is monotonically increasing in the price of its export good. Thus, the effect of discriminatory punishment on the deviating country depends on the effect of discrimination on that country's export prices. For example, take the case that country *X* has deviated from the agreement. Then the effect of discriminatory punishment can be reduced to the effect of lowered tariffs between the non-deviating countries (i.e.,  $\tau_{YZ}$ ,  $\tau_{ZY}$ ) on the export

prices of the deviating country (i.e.,  $p_X^y$ ,  $p_X^z$ ). For example, from the definition of  $p_X^y$ , one can derive that:

$$\frac{\mathrm{d}p_X^y}{\mathrm{d}\tau_{YZ}} = \frac{\partial p_X^y}{\partial \tau_{YZ}} + \frac{\partial p_X^y}{\partial \tau_{YX}} \cdot \frac{\partial \tau_{YX}}{\partial \tau_{YZ}}.$$
(14)

From (14) one sees that discriminatory punishment has two related effects on the welfare of the deviating country. First, since it allows the non-deviating countries to maintain lower trade barriers against one another, the basic impact of discrimination is to divert trade from the deviating country to the non-deviating countries. Thus, discriminating against the deviating country during periods of retaliation will directly lower the welfare of the deviating country by diverting trade away from that country. This can be seen in the first term of (14) where  $\partial p_x^{\nu} / \partial \tau_{YZ} > 0$ .

On the other hand, as discussed in Section 2, tariffs are complementary across countries (i.e.,  $\tau_{YX}/\partial \tau_{YZ} > 0$ ). When the non-deviating countries attempt to lower tariffs against one another, they are also lowering their unilaterally optimal tariff against the deviating country. Thus, nondiscriminatory punishment can credibly threaten higher tariffs against deviating countries than can a discriminatory punishment scheme. Accordingly, discriminating against the deviating country during periods of retaliation will indirectly raise the welfare of the deviating country since it will face lower retaliatory tariffs. This can be seen in the second term of (14) where  $\partial p_X^{\gamma}/\partial \tau_{YX} \cdot \partial \tau_{YZ} < 0$ .

As we establish in the following Lemma, this second effect is the stronger of the two and thus:

**Lemma 3.** Assuming that demand and supply functions are linear, non-discriminatory punishment results in lower welfare for the deviating country than discriminatory punishment in a period of retaliation  $(W_X^{NP} < W_X^{DP})$ .

**Proof.** From (14), (1) and (2) one can derive that lowering the tariff between non-deviating countries will raise the welfare of the deviating country if:

$$\frac{\partial \tau_{YX}}{\partial \tau_{YZ}} > \left(\frac{M_Z^{y'}}{M_Z^{y'} + M_Y^{y'}}\right) \tag{15}$$

Substituting (5), which provides an expression for  $\partial \tau_{YX} / \partial \tau_{YZ}$  into (15) we derive that  $\partial p_X^y / \partial \tau_{YZ} < 0$  if:

$$\left[M_{Y}^{y'} \cdot \sum_{J} M_{J}^{y'}\right] \frac{\partial \epsilon_{Z}^{y}}{\partial p_{Z}^{y}} > 0.$$
(16)

Since  $M_J^{i'} < 0$  for any good *i* and any country J and  $\partial \epsilon_Z^y / \partial p_Z^y > 0$  for linear demand and supply functions, the above condition is satisfied and a reduction in

the tariff between the nondeviating countries will raise the export prices and hence the welfare of the deviating country.  $\Box$ 

The result that tariff complementarity dominates trade diversion is due to the effect that trade diversion has on the trade elasticities (i.e.,  $\epsilon_{I}^{i}$ ). Intuitively, a reduction in the tariff between non-deviating countries (e.g.,  $\tau_{YZ}$ ) leads to increased trade between the non-deviating countries and reduced trade between the deviating country and the non-deviating countries. Focusing on the case where demand and supply functions are sufficiently linear, the reduced trade between the deviating country and non-deviating country leads to a lower optimal tariff against the non-deviating countries leads to a higher optimal tariff between the non-deviating countries leads to a higher optimal tariff between the non-deviating countries leads to a higher optimal tariff between the non-deviating countries (i.e.,  $\partial \epsilon_{Z}^{y} / \partial \tau_{YZ} < 0$ ) and thus, by (4), reduces the incentives to raise the tariff against the deviating country. Since both of these effects strengthen the tariff complementarity effect (by encouraging a lower tariff against the deviating country) it dominates the trade diversion effect and thus leads to non-discrimination being a superior punishment strategy.

However, if higher-order derivatives are relatively large, then the opposite result is possible. For example, if  $M_Z^{y''} = (M_Z^{y'})^2/M_Z^y$  then  $\partial \epsilon_Z^y/\partial \tau_Z^y = 0$  and, as can be seen from (16), the tariff-complementarity effect would equal the trade diversion effect and discrimination and nondiscrimination would be equivalent. By the same rationale, if the second derivatives dominate the first derivatives (i.e.,  $M_Z^{y''} < (M_Z^{y'})^2/M_Z^y < 0$ ) then  $\partial \epsilon_Z^y/\partial \tau_Z^y < 0$  and discrimination could actually be preferred. Intuitively, if the excess demand curve is sufficiently concave, then a reduction in the tariff between the non-deviating countries increases the elasticity of trade between the non-deviating countries thus weakening the tariff-complementarity effect so that it no longer dominates the trade diversion effect.

While the above qualification should be kept in mind and the limitations of the partial equilibrium model are acknowledged, nevertheless similar results are also present in general equilibrium models. For example, Kennan and Riezman (1990) show (in a general-equilibrium model of trade), that the creation of a free-trade agreement will shift the terms-of-trade in favor of and raise the welfare of non-member countries. The implication of this analysis is that maintaining cooperation among non-deviating countries will actually raise the welfare of the deviating country and thus lessen the threat of punishment. Thus, it is the case that within a standard model of self-enforcing agreements, the discriminatory power of trade policy could actually reduce the enforcement power of the agreement.

This result is reflected in Fig. 1, where, for a given T and  $\tau^c$ , since  $W_X^{NP} < W_X^{DP}$ , average welfare along the non-discriminatory punishment path is less than average welfare with discriminatory punishment. As can be seen in Fig. 1, the stronger

<sup>&</sup>lt;sup>9</sup>By sufficiently linear we mean that the first derivatives are large relative to the second derivatives.

punishment of non-discriminatory tariffs implies that the non-discriminatory agreement can support lower cooperative tariffs. Thus, from Lemma 2 and Lemma 3 we derive our first Proposition:

**Proposition 1.** Within a self-enforcing international agreement, holding T constant, non-discriminatory punishment will be preferred to discriminatory punishment when demand and supply functions are linear.

The above proposition implies that discriminating against cheating members in the punishment phase does not increase the severity of the punishment threatened to potential deviators. Surprisingly, discriminatory punishment actually lowers the punishment threat of the agreement, and thus non-discrimination would be the preferred punishment scheme. This result is a function of the fact that, when the non-deviating countries lower their bilateral trade barriers during periods of retaliation, they are providing a net terms-of-trade benefit to the deviating country.<sup>10</sup>

#### 4. Collusive punishment

In the previous section it was shown that allowing discriminatory tariffs during the punishment phase actually reduces the punishment threat of the agreement. However, this result was derived under the assumption that member countries (Y and Z) do not attempt to exploit their joint market-power against the deviating country (X) in the punishment phase. In this section we show that the opposite is possible if we allow member countries to collude against deviators by setting tariffs jointly (i.e., if countries Y and Z set tariffs against country X which maximize their joint-welfare during the punishment phase).

To see this, we first derive optimal (collusive) tariffs against the deviating country X under conditions of both discriminatory  $(\tau^{DC})$  and non-discriminatory  $(\tau^{NC})$  punishment. Under non-discriminatory punishment, countries Y and Z will set non-discriminatory tariffs against both country X and each other (i.e.,  $\tau_{YX} = \tau_{YZ} = \tau^{NC}$ ) to maximize their joint-welfare (i.e.,  $W_Y^y(\tau_{YX}, \tau_{YZ}) + W_Z^y(\tau_{YX}, \tau_{YZ})$ ). Taking the derivatives of this joint-welfare function with respect to the non-discriminatory trade barrier  $(\tau^{NC})$  and solving out the first-order condition one derives that:

<sup>&</sup>lt;sup>10</sup>The above analysis does assume that retaliation entails reversion to an interior Nash equilibrium. In the case that countries have access to export-sector policies and autarky is used as the Nash threatpoint, then discriminatory and non-discriminatory punishments are equivalent. Thus, in this section we can only state that non-discrimination is weakly preferred to discrimination as punishment scheme.

410 J. Ederington, P. McCalman / Journal of International Economics 61 (2003) 397-424

$$\tau_{YX}^{NC} = \tau_{YZ}^{NC} = \frac{M_X^y}{M_X^{y'} + M_Z^{y'}}$$
(17)

Note that the non-discriminatory collusive tariff  $(\tau^{NC})$  given in (17) is lower than the nondiscriminatory Nash tariff  $(\tau^N)$  given in (7). Intuitively, with collusion, MFN punishment tariffs are low because member countries are unwilling to raise tariffs against their colluding partners. Since these low tariffs decrease the punishment threat of the agreement, it should be apparent that when punishment is constrained to be non-discriminatory, the agreement should discourage collusion in the punishment phase.

With discriminatory punishment, tariffs set against country *X* are optimal collusive taxes for countries *Y* and *Z* (denoted  $\tau^{DC}$ ) conditional on the fact that countries *Y* and *Z* are continuing to cooperate among themselves (i.e.,  $\tau_{YZ} = \tau_{ZY} = \tilde{\tau}^c$ )). Once again, optimal collusive taxes are set to maximize the joint welfare of countries *Y* and *Z*. Taking the derivative of the joint-welfare function with respect to  $\tau_{YX}$  and solving out the first-order condition yields:

$$\tau_{YX}^{DC} = \epsilon_X^{y} + \tau_{YZ} \frac{M_Z^{y'}}{M_Y^{y'} + M_Z^{y'}}.$$
(18)

Note that for any  $\tau_{YZ} > 0$  the discriminatory collusive tariff ( $\tau^{DC}$ ) is greater than the unilaterally optimal tariff ( $\tau^{D}$ ) given in (4). Intuitively, with collusive taxes member countries can internalize the trade diversion effects of their tariffs. Thus, discriminatory tariffs will be high since member countries will be seeking to divert trade to their colluding partners, leading to lower welfare for deviating countries in the punishment phase of the agreement. As we show in the following proposition, this collusive effect will outweigh the tariff complementarity effect for small deviations from non-discriminatory Nash punishment (i.e., if  $\tilde{\tau}^{c}$  is sufficiently close to  $\tau^{N}$ ):

**Proposition 2.** For linear demand and supply functions, there exists a discriminatory punishment path, involving collusion among the non-deviating countries, which is preferred to non-discriminatory punishment in a self-enforcing agreement.

Proof. In Appendix B.

# 5. Renegotiation-proof agreements

In line with previous literature, the analysis of the previous section focused on trigger strategies in which violations are automatically punished for a fixed number of periods. However, one of the noteworthy features of the dispute settlement procedures (DSP) within the GATT system is the opportunities it provides for consultation and negotiation. Indeed the recent Uruguay round explicitly provides a forum in which member countries can discuss both perceived deviations and potential sanctions. As noted by Ludema (2001), if there is a mechanism through which countries can resolve their differences after deviation, then the potential severity of punishment may be limited. Specifically, once countries have reverted to the punishment phase, they will have an incentive to renegotiate so as to achieve some additional gains from trade. Unfortunately, such renegotiation reduces the threat of being punished and may undermine the previous cooperative agreement. The question raised in this section is whether the discriminatory power of tariffs provides a means of minimizing this potential problem.

To address this question, we focus our attention on international agreements that are 'renegotiation-proof'. There are many definitions of renegotiation proofness as well as punishment paths that satisfy certain definitions of renegotiation proofness. In this paper, we adopt the punishment path proposed by Ludema (2001) in which the time spent retaliating (defined in the previous section and denoted by T) is limited to ensure that the punishment is credible and is not susceptible to renegotiation. The advantage of adopting Ludema's proposed punishment scheme is that it is designed to mimic elements of the GATT dispute settlement procedure.

#### 5.1. Non-discriminatory punishment

In this section, we characterize a renegotiation-proof agreement that employs non-discriminatory retaliation. As before, a common cooperative tariff ( $\tau^c$ ) is negotiated in the initial phase to maximize cooperative welfare, subject to the provision that it is self-enforcing. In the event of a violation of the agreement, a punishment path is triggered. However, in this section, we are concerned with the possibility that, once countries have entered the punishment path, they will attempt to renegotiate the original agreement (and thereby avoid the threatened sanctions). To this end we follow Ludema (2001) by requiring that our punishment path has the following properties:

- (1) subgame perfection: no country has an incentive to unilaterally deviate;
- (2) reciprocity: each country receives equal welfare along the path;
- (3) consistency: the same path is specified for any deviation; and
- (4) unanimity: the path is Pareto efficient.

As shown by Ludema (2001), the above properties can be satisfied by a punishment path in which countries play symmetric Nash tariffs (i.e.,  $\tau^N$  defined by (7)) for *T* periods, followed by a return to the cooperative equilibrium (i.e, the lowest self-enforcing cooperative tariff). It should be apparent that the non-discriminatory punishment path specified above satisfies the first three properties. In addition, the fourth property can be satisfied by simply restricting the time spent

retaliating along the punishment path (T). Specifically, by adjusting T, the agreement can be configured so that the proposed punishment-path is pareto efficient (and thus will satisfy the unanimity requirement and be immune to renegotiation).

Note that for the agreement to be credible the proposed cooperative tariff  $(\hat{\tau}^c(\rho))$  must be self-enforcing. As a result, (12) must be satisfied to ensure that no country would deviate from the cooperative tariff.<sup>11</sup> From (12) one derives  $\hat{\tau}^c$  as a function of *T* (or equivalently,  $\rho \in [0,1]$ ). Note that the longer the punishment phase remains at the non-cooperative equilibrium (i.e., a lower  $\rho$ ), the more severe the punishment and the lower the  $\tau^c$  the agreement can support (i.e.,  $\partial \tau^c / \partial \rho > 0$ ). Thus, *T* must be sufficiently large so as to satisfy the self-enforcement constraint and support the proposed most-cooperative tariff.

The basic result of introducing renegotiation into the analysis is that the length of time spent retaliating is now endogenous. Specifically, a renegotiation-proof agreement will choose T (or  $\rho$ ) to maximize global welfare along the punishment path:

$$\max_{\rho} \omega^{NP} = (1 - \rho)W^{GNP} + \rho W^{GC}$$
(19)

where  $W^{GNP}$  represents global welfare in periods of non-discriminatory punishment (i.e.,  $W^{GNP} = W_X^{NP} + W_Y^{NP} + W_Z^{NP}$ ) and  $W^{GC}$  represents global welfare in periods of cooperation (i.e.,  $W^{GC} = W_X^C + W_Y^C + W_Z^C$ ). The solution to the above constrained maximization is demonstrated by Fig. 2. The payoff,  $W^C(\hat{\tau}^c)$ , is the cooperative level of welfare from a given 'most-cooperative' tariff. Thus,  $W^C$  is an increasing function of *T* (since greater punishment implies that a lower cooperative tariff can be supported). The average welfare along the punishment path ( $\omega$ ) is a weighted average of  $W^C$  and welfare in the punishment phase ( $W^{NP}$ ) and is equal to  $W^{NP}$  at T = 0 and at the limit as *T* approaches infinity. Thus, as can be seen in Fig. 2, by restricting the time spent in the punishment path. Solving (19) we find that the optimal  $\rho^*$  is given by the following first-order condition:)

$$\frac{\partial((1-\rho)W^{GNP})}{\partial\rho} + \frac{\partial(\rho W^{GC})}{\partial\rho} = 0$$
(20)

Of course reducing the time spent in the punishment phase limits the amount of cooperation that the agreement can achieve, but is necessary in order to make the required sanctions renegotiation-proof. Likewise, given  $\rho^*$ , the 'most-cooperative' tariff  $(\hat{\tau}^c(\rho))$  is given by the self-enforcement constraint (12). Assuming that the

<sup>&</sup>lt;sup>11</sup>There also exists a condition that the countries would be willing to enter the punishment phase (as opposed to reverting indefinitely to the Nash equilibrium) but this condition is satisfied trivially.



Fig. 2. Renegotiation - proof punishment.

self-enforcement constraint binds (and thus (12) holds with equality), by implicitly differentiating (12) one derives that at  $\hat{\tau}^c(\rho)$ 

$$\frac{\partial W_X^C}{\partial \rho} = (1-\delta) \frac{\partial W_X^D}{\partial \rho} + \delta \left[ \frac{\partial \left( (1-\rho) W_X^{NP} \right)}{\partial \rho} + \frac{\partial (\rho W_X^C)}{\partial \rho} \right]$$
(21)

Since non-discrimination results in symmetric welfare along the punishment path (i.e.,  $3W_X^C = W_C^G$  and  $3W_X^{NP} = W^{GNP}$ ), substituting (20) into (21) one derives that at the  $\hat{\tau}^c(\rho)$  for non-discriminatory punishment:

$$\frac{\partial W_X^C}{\partial \rho} = (1 - \delta) \frac{\partial W_X^D}{\partial \rho}.$$
(22)

As can be seen in Fig. 3, Eq. (22) implicitly defines the most-cooperative tariff



Fig. 3. Determining the most-cooperative tariff in a renegotiation - proof agreement.

for non-discriminatory punishment  $(\tau^{NP})$  as a function of  $\delta$  and  $\rho$ .<sup>12</sup> Note that as countries place higher weight on the future (i.e.,  $\delta \rightarrow 1$ ), the line  $(1 - \delta) dW^D$  shifts upward and the agreement can achieve a lower most-cooperative tariff (i.e.,  $\tau^{NP} \rightarrow 0$ ). The question we investigate in the following section is whether allowing discriminatory punishment will assist in achieving an even lower most-cooperative tariff.

<sup>&</sup>lt;sup>12</sup>Fig. 3 is drawn to reflect the fact that  $\partial W_x^C / \partial \rho = 0 > (1 - \delta) \partial W_x^D / \partial \rho$  at  $\tau^c = 0$  and that  $\partial W_x^C / \partial \rho < (1 - \delta) \partial W_x^D / \partial \rho$  at  $\tau^c = \tau^N$  for  $\delta \in (0,1)$  thus ensuring the existence of an equilibrium. Assuming well-behaved welfare functions (satisfying the second-order conditions that  $\partial W_x^C / \partial (\hat{\tau}^c)^2 < 0$  and  $\partial^2 \hat{\tau}^c / \partial \rho^2 < 0$ ) this equilibrium will be unique. If there are multiple equilibrium, we assume that countries coordinate to choose the lowest renegotiation-proof tariff and the analysis of this section goes through.

#### 5.2. Discriminatory punishment

In this section, we specify a punishment path that satisfies Ludema's properties for a renegotiation proof agreement with the exception that we modify the reciprocity condition to accommodate the possibility of discriminatory punishment. In Ludema (2001), reciprocity required that countries receive equal welfare from the punishment path, and was intended to mimic the GATT's emphasis on restoring a 'balance of concessions' in negotiations. With discriminatory punishment we maintain a balance of concessions since countries continue to set a common cooperative tariff ( $\tau^c$ ) in periods of cooperation, and deviators and non-deviators continue to play Nash tariffs against one another in periods of retaliation.

We adopt the same punishment path as before with the exception that, in a period of retaliation, the non-deviating countries maintain a lower tariff against each other ( $\tilde{\tau}^c$ ) than the Nash tariffs they play against the deviator. As we establish in the following Lemma, allowing non-deviating countries to maintain lower tariffs against each other ( $\tilde{\tau}^c$ ) in periods of retaliation means that the deviator experiences lower welfare along the punishment path than do the nondeviators:

**Lemma 4.** A small decrease in  $\tilde{\tau}^c$  from non-discriminatory punishment results in higher welfare for the non-deviating countries relative to the deviating countries.

## **Proof.** In Appendix C

We argue in this section that this relaxation of the reciprocity requirement can assist in solving the problem of renegotiation-proofness and thereby lead to a more cooperative agreement. Specifically, we argue that because the non-deviating countries are receiving higher welfare in periods of retaliation, discriminatory punishment can sustain punishment for a longer period of time, and thus can support a lower 'most-cooperative' tariff.

As before, to satisfy the unanimity condition, an agreement with discriminatory punishment must maximize global welfare along the punishment path:

$$\max_{\rho} \omega^{DP} = (1 - \rho)W^{GDP} + \rho W^{GC}$$
(23)

where  $W^{GDP}$  represents global welfare during a period of discriminatory retaliation (i.e.,  $W^{GDP} = W_X^{DP} + W_Y^{DP} + W_Z^{DP}$ ). In this section we maintain the assumption that country *X* is the deviating country, while countries *Y* and *Z* are the non-deviating countries.

The most-cooperative tariff  $(\hat{\tau}^c)$  set in periods of cooperation is determined by the self-enforcement condition for discriminatory punishment. Implicitly differentiating the self-enforcement constraint for discriminatory punishment (13), one derives that at  $\hat{\tau}^c(\rho)$ :

416 J. Ederington, P. McCalman / Journal of International Economics 61 (2003) 397-424

$$\frac{\partial W_X^C}{\partial \rho} = (1-\delta) \frac{\partial W_X^D}{\partial \rho} + \delta \left[ \frac{\partial ((1-\rho) W_X^{DP})}{\partial \rho} + \frac{\partial (\rho W_X^C)}{\partial \rho} \right].$$
(24)

Eq. (24) implicitly defines the most-cooperative tariff for discriminatory punishment ( $\tau^{DP}$ ) as a function of  $\delta$  and  $\rho$ . Note, from Fig. 3, that if discriminatory punishment does assist in achieving a lower most-cooperative tariff (i.e.,  $\tau^{DP} < \tau^{NP}$ ), then it must be the case that  $\tau^{DP}$  is in the area where  $\partial W_X^C / \partial \rho > (1 - \delta) \partial W_X^D / \partial \rho$ . Thus, we can derive Lemma 5:

**Lemma 5.** Within a renegotiation-proof agreement, discriminatory punishment assists in achieving a lower 'most-cooperative' trade barrier if:

$$\frac{1}{3}W^{GDP} - W_X^{DP} + (1-\rho) \left[ \frac{\partial W_X^{DP}}{\partial \tilde{\tau}^c} - \frac{1}{3} \frac{\partial W^{GDP}}{\partial \tilde{\tau}^c} \right] \frac{\partial \tilde{\tau}^c}{\partial \rho} > 0.$$
(25)

Proof. From (24), if

$$\frac{\partial [(1-\rho)W_X^{DP}]}{\partial \rho} + \frac{\partial [\rho W_X^C]}{\partial \rho} > 0 \quad \text{then} \quad \frac{\partial W_X^C}{\partial \rho} > (1-\delta) \frac{\partial W_X^L}{\partial \rho}$$

at  $\tau^{DP}$ . As can be seen from Fig. 3, this implies that  $\tau^{DP} < \tau^{NP}$ . It is direct to derive that:

$$\delta \left[ \frac{\partial ((1-\rho)W_X^{DP})}{\partial \rho} + \frac{\partial (\rho W_X^C)}{\partial \rho} \right] = W_X^C - W_X^{DP} + \rho \frac{\partial W_X^C}{\partial \rho} + (1-\rho) \frac{\partial W_X^{DP}}{\partial \rho}$$
(26)

Likewise, from (23), the first-order condition for the optimal  $\rho^*$  for discriminatory punishment reduces to:

$$W^{GC} - W^{GDP} + \rho \frac{\partial W^{GC}}{\partial \rho} + (1 - \rho) \frac{\partial W^{GDP}}{\partial \rho} = 0$$
(27)

Since discrimination results in symmetric welfare only during the cooperative phase of the agreement (i.e.,  $W^{GC} = 3W_X^C$ ), one can substitute (27) into (26) to derive that  $\delta[\partial(1-\rho)W_X^{DP}/\partial\rho + \partial\rho W_X^C/\partial\rho] > 0$  and thus  $\tau^{DP} < \tau^{NP}$  if (25) is satisfied.  $\Box$ 

According to Lemma 5, if (25) is satisfied, then discriminatory punishment is beneficial in minimizing the problems of renegotiation and achieving a lower most-cooperative tariff. Note that the sole difference between discriminatory and non-discriminatory punishment is that nondeviating countries play Nash tariffs against each other with non-discrimination (i.e.,  $\tau_{YZ} = \tau_{ZY} = \tau^N$ ) while maintaining some degree of cooperation with discriminatory punishment (i.e.,  $\tau_{YZ} = \tau_{ZY} = \tilde{\tau}^c$ ).

Assume that the tariff set between non-deviating countries ( $\tilde{\tau}^c$ ) is given exogenously, and is independent of the time spent retaliating  $\rho$ . Under this assumption, welfare during periods of retaliation is independent of  $\rho$  and (25) reduces to the condition that:

$$\frac{1}{3}W^{GDP} > W_X^{DP} \tag{28}$$

Using (28), Lemma 5 states that discriminatory punishment will be beneficial in supporting a lower most-cooperative tariff if welfare for the deviating country is less than average welfare along the punishment path. However, we have already established in Lemma 4 that this is the case for  $\tilde{\tau}^c$  close to  $\tau^N$ . Thus, we can state Proposition 3:

**Proposition 3.** A discriminatory punishment path exists which will be preferred to nondiscriminatory punishment in a renegotiation-proof agreement.

Proposition 3 is based on the fact that a small movement towards discriminatory punishment will benefit the non-deviating countries more than the deviating country. Even though average welfare along the punishment path may be the same with both discriminatory and nondiscriminatory punishment (since the higher welfare in periods of retaliation with discriminatory punishment is balanced by a longer amount of time spent retaliating), the deviating country will receive less than average welfare and thus more severe punishment with discriminatory punishment.

However, not all discriminatory schemes are superior to non-discrimination. It is possible for certain values of  $\tilde{\tau}^c < \tau^N$ , that welfare for the deviating country is higher than average global welfare along the punishment path. To show this, we parameterize our model by assuming that demand for each country *J* on each good *i* is given by  $D_J^i(p_J^i) = a - b \cdot p_J^i$ . Similarly, supply for each country on its import good is given by  $Q_J^i(p_J^i) = c \cdot p_J^j$ , while the supply functions for the export goods are given by  $Q_J^i(p_J^i) = d \cdot p_J^j$  (with d > c). Solving out the parameterized model and plotting welfare for the deviating country versus average welfare along the punishment phase, one derives Fig. 4.<sup>13</sup> As can be seen, while small deviations from the Nash equilibrium result in relatively lower welfare for the deviating country, it is possible that large deviations from non-discrimination will result in the deviating country receiving higher than average welfare in the punishment

<sup>&</sup>lt;sup>13</sup>Fig. 4 is drawn to correspond with the parameter values b = c = 1 and d = 2. If either b or c is sufficiently large, or d is sufficiently small, then  $\tilde{\tau} < 0$  and thus discrimination would be preferred for the entire range of  $\tilde{\tau}^c \in [0, \tau^N]$ .



Fig. 4. Discrimination vs. non-discrimination with renegotiation.

path.<sup>14</sup> Thus, for  $\tilde{\tau}^c < \tilde{\tau}$  in Fig. 4, non-discriminatory punishment will be preferred in a renegotiation-proof agreement.

In Fig. 4,  $\tilde{\tau}$  divides  $\tilde{\tau}^c$  values into a region where discrimination would be preferred and a region where non-discrimination would be preferred. The fraction  $\tilde{\tau}/\tau^N$  represents the proportion of  $\tilde{\tau}^c \in [0, \tau^N]$  such that non-discrimination is preferred. Solving out for  $\tilde{\tau}$  and  $\tau^N$  in terms of parameter values, it is direct to derive that  $\tilde{\tau}/\tau^N$  is monotonically increasing in *d* and decreasing in *c* and *b* (for any positive values of these parameters where d > c). This is due to the fact the tariff-complementarity effect (given by (6)) is also monotonically increasing in *d* and decreasing in *c* and *b*. Intuitively, for certain parameter values, the tariff complementarity effect is very large and the deviating country is more likely to

<sup>&</sup>lt;sup>14</sup>This counter-intuitive result can also be seen, for certain parameter values, in Kennan and Riezman (1990) where the country outside the free-trade agreement actually receives higher welfare than the countries that have signed the free-trade agreement.

benefit from a large reduction in the tariff between non-deviating countries. Thus, non-discrimination is more likely to be preferred when comparative advantage (i.e., d-c) is large since a greater comparative advantage implies a larger tariff-complementarity effect. Likewise, discrimination is more likely to be preferred when the demand curve is steep (i.e., b is large) since a steep demand curve implies a smaller tariff-complementarity effect.

Another qualification to Proposition 3 is the possibility that  $\tilde{\tau}^c$  will be a function of  $\rho$ . For example, assume that the non-deviating countries simply maintain the most-cooperative tariff in periods of retaliation (i.e.,  $\tilde{\tau}^c = \hat{\tau}^c(\rho)$ ). In this case, discriminatory punishment will be preferred only if (25) is satisfied (where  $\partial \tilde{\tau}^c / \partial \rho > 0$ ). This assumption will actually have the effect of expanding the parameter values for which non-discrimination would be preferred, since (as can be seen in Fig. 4) there is a wide range of  $\tilde{\tau}^c$  for which

$$\frac{\partial W_X^{DP}}{\partial \tilde{\tau}^c} - \frac{1}{3} \frac{\partial W^{GDP}}{\partial \tilde{\tau}^c} < 0.$$

However, given Lemma 4, it is still the case that a small movement away from non-discriminatory punishment will be beneficial as

$$\frac{\partial W_X^{DP}}{\partial \tilde{\tau}^c} - \frac{1}{3} \frac{\partial W^{GDP}}{\partial \tilde{\tau}^c} > 0$$

in the area around the Nash equilibrium (i.e., when  $\tilde{\tau}^c = \hat{\tau}^c(\rho)$  is close to  $\tau^N$ ).<sup>15</sup>

#### 6. Conclusion

In this paper we are concerned with the question of whether exceptions to the MFN principle should be made within the dispute phase of international agreements. We show that, within a standard model of a self-enforcing agreement, discriminatory punishment actually reduces the severity of punishment that can be threatened to potential cheaters, and thus leads to a less cooperative agreement. Therefore, countries would be better off if they could credibly commit to non-discriminatory punishment (symmetric punishment of all members of the agreement) since such punishment is a stronger deterrent to deviations.

However, we also derive two cases under which discriminatory punishment is beneficial. First, discriminatory punishment is beneficial when member countries can collude against cheaters in the punishment phase. Second, discriminatory punishment is beneficial since it minimizes the enforcement problems created by

<sup>&</sup>lt;sup>15</sup>Note that, when  $\tilde{\tau}^c = \hat{\tau}^c(\rho, \delta)$ , the relative benefits of discrimination will be tied to the discount factor as, when countries discount the future heavily,  $\hat{\tau}^c(\rho, \delta)$  will be close to the Nash equilibrium and thus discrimination will be preferred.

the potential for renegotiation and thereby allows the agreement to credibly threaten more stringent punishment against deviating countries. Our results imply that discriminatory trade barriers may be the preferred instrument of enforcement in those agreements that are susceptible to the problems of renegotiation.

The results also have important implications for the use of trade sanctions as a means of enforcing agreements covering domestic policies. The success of GATT negotiations in lowering tariff barriers has turned attention to the use of domestic policies as secondary instruments of protection. Indeed, many of the current trade disputes, whether it is the US challenging current liquor taxes in Chile, or Venezuela challenging the system of gasoline taxes in the US, revolve around domestic policies (i.e., internal taxes and regulations). Cooperation over domestic policies will only gain in importance in the coming years as issues involving labor and environmental standards move to the forefront. However, as international negotiations concentrate on a country's domestic policies, one unanswered question remains the proper means of enforcement. Specifically, should 'side' agreements over a country's domestic policies be enforced with the threat of the suspension of trade concessions? This question is especially contentious given that current GATT/WTO rules forbid the use of trade sanctions as a threat to either induce members to accept obligations or to enforce international cooperation in agreements outside of those negotiated within GATT/WTO.

Previous theoretical research on the question of issue linkage has analyzed it within bilateral models of cooperation. For example, Cesar and de Zeeuw (1996) show that given asymmetry across the two countries, issue linkage may be optimal. However, Abrego et al. (2001) argue that side payments are a more efficient means of ensuring cooperation among asymmetric countries. In addition, Ederington (2002) shows that the stronger threat of trade policy sanctions is not necessary to enforce an agreement covering domestic policies. Thus, previous research using bilateral models of cooperation has failed to make a strong case for issue linkage (or the use of trade policy sanctions as a means of enforcing agreements covering domestic policies). However, such models ignore the fact that one of the inherent advantages of trade policy over domestic policy as an enforcement mechanism is its discriminatory potential. While a general consumption tax on wheat does not distinguish between US, Russian or Australian wheat, a country-specific tariff does. One of the advantages of trade sanctions may lie in its ability to punish only those members who are cheating on the agreement, while maintaining cooperation among the rest.

For example, assume the three-country model of Section 2, but now let each country have access to two policy instruments: trade policy and domestic policy. Domestic policy represents all the internal rules and regulations of the importing country that apply to the sale and production of the good while trade policy represents taxes placed upon the good on arrival in the country and can be specific to the nation of origin. Assume that countries can craft a domestic policy that is a perfect substitute for a non-discriminatory tariff (e.g., a combination consumption

tax and production subsidy). Then the sole difference between trade policy and domestic policy is that trade policy can be discriminatory. It should then be immediately apparent that all the results of our paper are applicable to the use of trade sanctions as a means of enforcing international agreements covering nondiscriminatory domestic policies. For example, assume that countries form an agreement to set cooperative domestic policies. Under the standard self-enforcing model of Section 3, the optimal punishment will entail a non-discriminatory reversion to the Nash equilibrium. This punishment can be accomplished by a reversion to the Nash equilibrium in domestic policies alone, and thus trade sanctions will not assist in maintaining greater cooperation. However, if punishments are required to be renegotiation-proof (as in Section 5) then the optimal punishment will entail discrimination against the deviating country. Since such discrimination cannot be accomplished through domestic policy, the discriminatory potential of trade sanctions will assist in maintaining greater cooperation. Thus, the results of our paper suggest that trade sanctions may be a useful enforcement mechanism for those agreements that are sensitive to the problems of renegotiation.

#### Acknowledgements

We would like to thank Robert Staiger and an anonymous referee for very helpful comments. We would also like to thank Eric Bond, Steven Matusz and seminar participants at both the 1999 Southeast International Economics Meetings and the Spring 2000 Midwest International Economics Meetings.

# Appendix A. Proof of Lemma 2

Rearranging (12) one derives that a cooperative tariff  $(\tau^c)$  is supported within the non-discriminatory agreement provided that  $W^{NP} \leq \phi(\tau^c)$  where  $\phi(\tau^c) \equiv W^C - [(1-\delta)/\delta](W^D - W^C)$ . From the definitions of  $W^C$  and  $W^D$  one can derive that  $\partial \phi/\partial \tau^c \geq 0$  at  $\tau^c = 0$  and  $\partial \phi/\partial \tau^c \leq 0$  at  $\tau^c = \tau^N$ . Likewise,  $\partial [(1-\rho)W^{NP} + \rho W^C]/\partial \tau^c \leq 0$  for all  $\tau^c \in [0, \tau^N]$ . Finally, note that  $\phi = [(1-\rho)W^{NP} + \rho W^C]$ and  $\partial \phi/\partial \tau^c \leq \partial [(1-\rho)W^{NP} + \rho W^C]/\partial \tau^c$  at  $\tau^c = \tau^N$ . Therefore, for a sufficiently low discount factor, the self-enforcement constraint will bind and there exists a  $\tau^{NP} \in [0, \tau^N]$  such that  $[(1-\rho)W^{NP} + \rho W^C] = \phi(\tau^{NP})$ , and  $[(1-\rho)W^{NP} + \rho W^C] < \phi(\tau^c)$  for  $\tau^c \geq \tau^{NP}$ . This set-up is reflected in Fig. 1 where  $\tau^{NP}$  is the lowest cooperative tariff for which the self-enforcement constraint binds. It should then be apparent that, if  $W^{DP} \geq W^{NP}$ , then  $\tau^{DP} \geq \tau^{NP}$  (where  $\tau^{DP}$  is the lowest cooperative tariff for which the self-enforcement constraint binds with discriminatory punishment). Likewise, if  $W^{DP} \leq W^{NP}$ , then  $\tau^{DP} \leq \tau^{NP}$ .  $\Box$ 

# Appendix B. Proof of Proposition 2

From (7) and (18) one can compare the collusive tariff with discrimination  $(\tau_{YX}^{DC})$  to the noncollusive, non-discriminatory tariff  $(\tau_{YX}^{N})$ :

$$\tau_{YX}^{DC} - \tau_{YX}^{N} = \left[ \boldsymbol{\epsilon}_{X}^{y} (\tau_{YX}^{DC}, \tilde{\tau}^{c}) - \boldsymbol{\epsilon}_{X}^{y} (\tau_{YX}^{N}, \tau_{YZ}^{N}) \right] + \tilde{\tau}^{c} \frac{M_{Z}^{y}}{M_{Y}^{y'} + M_{Z}^{y'}}$$
(B.1)

Employing a Taylor series expansion of the term  $\left[\epsilon_{X}^{y}(\tau_{YX}^{DC}, \tilde{\tau}^{c}) - \epsilon_{X}^{y}(\tau_{YX}^{N}, \tau_{YZ}^{N})\right]$  and assuming linear supply and demand functions one can simplify (B.1) to:

$$\tau_{YX}^{DC} - \tau_{YX}^{N} = \left[ \tilde{\tau}^{c} - \tau_{YZ}^{N} \frac{M_{Y}^{y'} + M_{Z}^{y'}}{\sum_{J} M_{J}^{y'} + M_{Y}^{y'} + M_{Z}^{y'}} \right] \frac{M_{Z}^{y'}}{M_{Y}^{y'} + M_{Z}^{y'}}$$
(B.2)

Thus, from (B.2) one can derive that the discriminatory tariff with collusion is greater than the non-discriminatory tariff  $(\tau_{YX}^{DC} > \tau_{YX}^{N})$  provided that  $\tilde{\tau}^{c}$  is sufficiently close to the Nash tariff. Given that  $\tilde{\tau}^{c} < \tau_{YZ}^{N}$  this is a sufficient condition for the deviating country (X) to receive lower welfare along the punishment path.

# Appendix C. Proof of Lemma 4

The effect of a small decrease in the tariff between non-deviating countries  $(\tau_{YZ} = \tilde{\tau}^c)$  on the welfare of the deviating country is given by:

$$\frac{\partial W_X^y}{\partial \tilde{\tau}^c} = -M_X^y \frac{\partial p_X^y}{\partial \tilde{\tau}^c} \tag{C.1}$$

Likewise, the effect on global welfare of a small decrease in  $\tilde{\tau}^c$  (evaluated at the Nash equilibrium) is given by:

$$\frac{\partial (W_X^y + W_Z^y + W_Y^y)}{\partial \tilde{\tau}^c} = -M_X^y \frac{\partial p_X^y}{\partial \tilde{\tau}^c} - M_Z^y \frac{\partial p_Z^y}{\partial \tilde{\tau}^c}$$
(C.2)

Given the symmetry of the model (i.e.,  $M_Z^y = M_X^y$  at the Nash equilibrium) and using (1) and (2), one can derive that

$$\frac{1}{3} \frac{\partial (W_X^y + W_Z^y + W_Y^y)}{\partial \tau^c} < \frac{\partial W_X^y}{\partial \tau^c} < 0$$

if:

$$\frac{\partial \tau_{YX}^{D}}{\partial \tilde{\tau}^{c}} < \frac{M_{Z}^{y'} + \sum_{j} M_{J}^{x'}}{M_{Z}^{y'} + M_{y}^{y'} + \sum_{j} M_{J}^{x'}}$$
(C.3)

Thus, according to (C.3), a decrease in  $\tilde{\tau}^c$  from the Nash equilibrium will benefit the nondeviating countries more than the deviating country if the tariff complementarity effect is sufficiently small. Using (6), one can show that (C.3) is satisfied if:

$$\sum_{j} M_{J}^{x'} \left[ M_{Y}^{y'} \left( 2M_{Z}^{y'} + M_{Z}^{y'} \right) \right] < 0$$
(C.4)

Since  $M_J^{i'} < 0$  for each good *i* and country *J*, it is direct to derive that the above condition is satisfied.  $\Box$ 

# References

- Abrego, L. et al., 2001. Trade and environment: bargaining outcomes from linked negotiations. Review of International Economics 9, 414–428.
- Bagwell, K., Staiger, R., 1990. A theory of managed trade. American Economic Review 80, 779-795.
- Bagwell, K., Staiger, R., 1999a. An economic theory of GATT. American Economic Review 89, 215–248.
- Bagwell, K., Staiger, R., 1999b. Regionalism and multilateral tariff co-operation. In: Piggott, J., Woodland, A. (Eds.), International Trade Policy and the Pacific Rim. St. Martin's Press.

Bagwell, K., Staiger, R., 2002. Multilateral Trade Negotiations, Bilateral Opportunism and the Rules of GATT/WTO. Journal of International Economics (in press).

- Bond, E. et al., 2001. Deepening of regional integration and multilateral trade agreements. Journal of International Economics 53, 335–361.
- Cesar, H., de Zeeuw, A., 1996. Issue linkage in global environmental problems. In: Xepapadeas, A. (Ed.), Economic Policy for the Environmental and Natural Resources. Edward Elgar Press.
- Dixit, A., 1987. Strategic aspects of trade policy. In: Bewley, T. (Ed.), Advances in Economic Theory: Fifth World Congress. Cambridge University Press, Cambridge.
- Ederington, J., 2002. Trade and domestic policy linkage in international agreements. International Economic Review 43, 1347–1367.
- Hungerford, T., 1991. GATT: A cooperative equilibrium in a non-cooperative trading regime. Journal of International Economics 31, 357–369.
- Jackson, J., 1989. In: The World Trading System: Law and Policy of International Economic Relations. MIT Press, Cambridge, MA.
- Kennan, J., Riezman, R., 1990. Optimum tariff equilibria with customs unions. Canadian Journal of Economics 23, 70–83.
- Ludema, R., 2001. Optimal international trade agreements and dispute settlement procedures. European Journal of Political Economy 72, 355–376.
- Maggi, G., 1999. The role of multilateral institutions in international trade cooperation. American Economic Review 89, 190–214.

# 424 J. Ederington, P. McCalman / Journal of International Economics 61 (2003) 397-424

- McCalman, P., 2002. Most-favored nation clause and multilateral trade negotiations. Journal of International Economics 57, 151–176.
- Riezman, R., 1991. Dynamic tariffs with asymmetric information. Journal of International Economics 30, 267–283.

Zissimos, B., Vines, D., 1999. Is Article XXIV Bad? manuscript.