

A Prisoner's Dilemma Tariff Setting Game with an Escape Clause

Mathias Herzing^{*†}

Abstract

This paper attempts to provide a game theoretic rationale for the inclusion of escape clause mechanisms in trade agreements, by examining the effects of uncertain outcomes on the sustainability of cooperation in an infinitely repeated Prisoner's Dilemma tariff setting game. Since, in the presence of uncertainty affecting the outcome under cooperation, deviations and thus the breakdown of cooperation may occur, it is necessary to design an agreement such that cooperation can be sustained under all contingencies. The introduction of an escape clause that allows a shock-affected country to temporarily deviate, while sufficiently compensating its trading partner in order to alleviate the loss from being exposed to the deviation, offers such a possibility. It is shown that with such a mechanism cooperative behavior can always be sustained. Moreover, the expected per-period payoff increases as compared to the case when there is no escape clause. For a sufficiently high discount factor, the threshold level for temporary deviation through the escape clause will be at the efficient level. It is shown that, when the optimal compensation cost can be applied, the optimal cooperative is strictly smaller than the optimal cooperative tariff in the absence of uncertainty. Furthermore it is shown that a decrease in the cooperative tariff unambiguously leads to an increase in the use of the escape clause.

^{*}Institute for International Economic Studies, Stockholm University, August 2003

[†]This paper is a draft of what will be the first essay of my dissertation. I am grateful to my advisors Harry Flam and Henrik Horn. I would also like to thank the participants of the Stockholm Trade Workshop in September 2002 for valuable comments. Financial support from the Tom Hedelius and Jan Wallander Research Foundation is gratefully acknowledged.

1 Introduction

The two main features of post-war trade have been on the one hand a lowering of tariff levels, and on the other hand an increasing use of non-tariff barriers (NTBs). As liberalization has continued and trade has expanded, countries have become increasingly exposed to foreign markets and the world market. This increasing exposure to trade flows, while having an overall beneficial effect through the gains from trade, has created increased incentives for countries to temporarily deviate from cooperation in order to respond to political economic strains due to unforeseen circumstances. Having committed to lower tariff levels, countries have therefore increasingly made use of various NTBs in response to domestic political economic pressures. Since the increasing use of such trade distorting measures have threatened to undermine the achievements of trade liberalization, trade negotiations have come to focus on removing the NTBs and finding mechanisms allowing countries to temporarily deviate from the cooperative regime under certain circumstances. The underlying rationale for such negotiations is that, as liberalization and hence exposure to trade flows increases, countries that face unanticipated, supposedly trade-related, shocks will need to be allowed to take temporary measures without thereby threatening overall liberalization. It is also considered likely that, by allowing for exceptions after a trade agreement has been implemented, the degree of cooperation will increase and the chances for reaching an agreement in the first place are increased.

During the past two decades a great variety of game theoretic models have been developed in order to explain the trade political setting within which countries operate. They can be divided into three broad categories:

- Cooperative games with some (implicitly assumed) enforcement mechanism.¹
- Non-cooperative games without an enforcement mechanism.²
- Non-cooperative games with an (explicit) enforcement mechanism.³

Since the beginning of the 1980s there has been a continuing process starting out from static descriptions of trade policy in cooperative

¹See, for example, Mayer (1981) and Riezman (1982).

²See, for example, Dixit (1987), Bagwell and Staiger (1990), and Riezman (1991).

³See, for example, Hungerford (1991), Kovenock and Thursby (1992), Maggi (1999), and Ludema (2001).

settings to increasingly sophisticated attempts to model enforcement in non-cooperative settings. During the same period the Uruguay Round was completed and the World Trade Organization (WTO) was founded, evolving from the General Agreement on Tariffs and Trade (GATT) and including a reformed Dispute Settlement Procedure (DSP). Several models attempting to incorporate elements of the vast variety of different features of the WTO and its DSP (e.g. the Most-Favored-Nation clause, reciprocity, concession diversion, and commensurate punishment) have appeared recently, e.g. Bagwell and Staiger (2001) and Ethier (2001).

The model used here draws on an approach used in Rosendorff and Milner (2001) who apply a two-stage game between two countries. At the first stage, in an international bargaining game, negotiations over the design of the institutional framework take place. At the second stage, an infinitely repeated trade policy game between countries, given the design of the institution, takes place. In each period the political pressure for protection at home (and/or for more open markets abroad) is subject to a shock. This shock can be seen as any exogenous and unanticipated change in the state of the world (unexpected price or supply shifts, changes in production technology, changes in a country's political institutions or preferences, changes in domestic political cleavages or alignments) that affects domestic firms' demand for, or ability to lobby for, protection of their markets. It is furthermore assumed that the two countries' shocks are stochastic and independently and identically distributed, that in the current period each country knows its own state of politics but not the other's, and that both are equally uninformed about the weight values (at home and abroad) in all future periods.

In each period the countries find themselves in a Prisoner's dilemma. By deviating, a short-term gain can be made, but at the cost of infinite Nash reversion thereafter (given that grim-trigger strategies are applied). Ex ante, expectations are formed about the outcomes under mutual cooperation, deviation, being deviated against and the Nash equilibrium. It is assumed that there is an upper bound to the short-term gain that can be made by deviating, and it is shown that for a sufficiently high discount factor cooperation forever can be sustained. However, since cooperation may break down at lower discount factors, the possibility of exercising an escape clause at a fixed cost, allowing deviation for one period, is introduced. Having used the escape clause for one period the country returns to the cooperative regime in the next period, having preserved its reputation as a cooperator. No supranational enforcement agency is necessary to make the escape-clause-using country pay the cost, because it wishes to preserve its credibility in the future. The institu-

tion merely serves as a verification agency, much as the Law Merchants institution did.⁴

It is shown that in the equilibrium with an escape clause, deviation can be avoided for any discount factor, as long as this cost is not larger than a threshold value that is increasing in the discount factor. Moreover, it is established that either there is an escape clause with a level of cost that induces enough cooperation and no breakdown such that the value of the game in an escape clause equilibrium is larger than that of the same game without an escape clause, or the cost of escape is too high and the escape clause equilibrium is the same as the grim-trigger equilibrium in the absence of an escape clause. The model predicts that greater domestic uncertainty, or situations where political leaders are more sensitive to unanticipated changes in political pressures, should be associated with more reliance on escape clauses. Hence, countries that are more sensitive to domestic pressures should be the main proponents and users of escape clauses, and certain issue-areas, such as exchange rate mechanisms and trade agreements, should be more likely to have escape clauses than others due to their greater levels of uncertainty.

Three central conclusions are drawn from the model. First, when political leaders cannot foresee the extent of future domestic demands for more protection at home (and/or more open markets abroad), escape clauses provide the flexibility that allows them to accept an international trade liberalizing agreement. The flexibility provided in an agreement, in terms of the ability to adapt and respond to unanticipated events within the context of a well-designed institutional system, increases with uncertainty. Second, for escape clauses to be useful and efficient they must impose some kind of cost on their use, in order to signal the intention of its user to comply in future. Thus an agreement will be designed such that the costs of the escape clauses they most desire are balanced by the benefits of future cooperation. Third, including escape clauses make initial agreements easier to reach. Increased flexibility, in order to deal with the uncertainty about future states of the world, lessens the problems of bargaining and distribution that may inhibit an initial agreement. Thus flexibility increases with distribution problems.

This paper also starts out from the traditional game theoretic approach, modeling trade interaction as an infinitely repeated Prisoner's Dilemma game. In the absence of cooperation countries apply their respective optimal tariffs vis-à-vis each other and hence are stuck in a

⁴See Milgrom, North, and Weingast (1990).

suboptimal Nash equilibrium. Cooperation in this setting can only be achieved and sustained if there is strong enough punishment against a deviator, e.g. infinite reversion to the Nash equilibrium. If however shocks that influence the incentive to deviate from cooperation occur, cooperation will break down if these are strong enough. Following Rosendorff and Milner (2001), in order to deal with such unanticipated shocks (which here are assumed to be unbounded) and avoid the breakdown of cooperation, an escape clause, allowing a shock-affected country to temporarily deviate from cooperation without causing infinite reversion to the Nash outcome, can be included in an agreement. By making this modification to the game, negotiations will thus be as follows. Two symmetric countries negotiate a trade agreement consisting of two elements, a cooperative tariff level and principles governing behavior in response to unforeseen events. At the first stage, there are negotiations over the rules prescribing what action should be allowed in response to an unanticipated shock that decreases the payoff under cooperation vis-à-vis deviation. For tractability it is assumed that a shock-affected country will apply its optimal tariff and at the same time incur some cost for its temporary deviation. Different from Rosendorff and Milner (2001) this cost is not sunk but translates into a benefit for the trading partner, thus alleviating the damage of being exposed to the optimal tariff by the other country. At the second stage, after principles for compensatory measures have been agreed upon, the level of cooperation subject to these principles is negotiated.

The model presented here is related to theories of infinitely repeated oligopoly games in the presence of unobservable exogenous shocks. Green and Porter (1984) investigate how collusion in a Cournot oligopoly is affected by unobservable demand shocks that influence the price level. In this model a single firm will not know whether a low market price is due to cheating by a competitor or low demand. By prespecifying a certain price level below which reversion to the Cournot Nash equilibrium will take place, it is possible to achieve collusive behavior. Episodes of too low prices, followed by Cournot Nash reversion, will nevertheless occur, in order to sustain the agreement. Abreu, Pearce and Stacchetti (1986) modify and generalize the Green-Porter model and show that in equilibrium only two quantities are ever produced. It is also shown that a firm simply needs to remember the previous period's price and what quantity was specified by the equilibrium in that previous period. While the model presented here also attempts to capture the effect of uncertainty on cooperation, there is a fundamental difference compared to the Green and Porter setting: Here each country will experience a shock that it observes perfectly, whereas its trading partner cannot. The action it takes,

i.e. either cooperate or deviate, will become common knowledge after its implementation, contrary to the Green and Porter model. In this setting it is thus not a matter of trying to conceal and detect protectionist behavior.

In the next section the various safeguard actions permitted under the legal framework of the GATT and the WTO, with special emphasis on Article XIX of the GATT, are presented. Section three examines the case of the infinite repetition of a Prisoner’s Dilemma game in the presence of unanticipated shocks. Then, in section four, an escape clause is introduced and its implications are investigated. The optimal design of an escape clause mechanism in a special case is presented in the fifth section. Section six concludes.

2 The Legal Framework

2.1 The Safeguard Provisions of the GATT

The term “escape clause” usually refers to the safeguard measures permitted under Article XIX of the GATT. More generally, a safeguard is a provision that allows a WTO member to withdraw or cease to apply its normal obligations in order to protect certain overriding interests under specified conditions. Safeguard provisions on the one hand provide governments with the means to deviate from specific liberalization commitments in certain circumstances and thus have a safety-valve function, and on the other hand facilitate the signing of a protection-reducing agreement and thus serve as an insurance mechanism (Hoekman and Kostecki, 1995).

It is possible to distinguish the various safeguard measures permitted under the GATT with regard to whether they are temporary or permanent and to whether they are targeted at a specific industry or have economy-wide implications (Hoekman and Kostecki, 1995). The following table contains the safeguard provisions under the GATT, separated according to scope and duration.

	Temporary action	Permanent action
Industry-specific	Article VI Article XVIII: A and C Article XIX Article XXV	Article XXVIII
Economy-wide	Article XII Article XVIII: B	Article XX Article XXI

The various objectives of these safeguard provisions are summarized as follows in Hoekman and Kostecki (1995).

- Combating “unfair” trade by applying antidumping and countervailing duties (Article VI)
- Establishment of an industry, i.e. infant-industry protection (Article XVIII Sections A and C)
- Facilitating adjustment of an industry, i.e. emergency protection (Article XIX)
- Seeking a derogation (waiver) from specific GATT rules (Article XXV)
- Alleviating balance of payment problems (Articles XII and XVIII Section B)
- Allowing for renegotiation of tariff concessions (Article XXVIII)
- Achieving health, safety and related objectives (Article XX)
- Maintaining national security (Article XXI)

With regard to temporary and unexpected pressures stemming from sudden increases in imports, the GATT thus contains a range of measures for signatory countries that can be applied in order to alleviate these. A WTO member may restrict imports of a product temporarily in order to protect a specific domestic industry from an increase in imports that is causing, or threatening to cause, serious injury to the industry. There are several articles in the GATT that can be invoked to that purpose. First, Article VI contains provisions for dealing with dumping and export subsidization that harms the industry of the importing country or of another exporting country. Countries are allowed to impose antidumping duties in order to alleviate the damage of dumping, and countervailing duties are permitted to offset “any bounty and subsidy bestowed, directly, or indirectly, upon the manufacture, production or export of any merchandise” (Art VI §3). Second, Article XII allows a contracting party to restrict imports “in order to safeguard its external financial position and its balance of payments” (Art XII §1). Third, Article XVIII addresses the potential need for developing countries to take protective measures. Especially the promotion of “the establishment of particular industries with a view to raising the general standard of living of its people” (Art XVIII §3) is emphasized. Fourth, Article XIX provides the possibility

of taking emergency actions on imports of particular products. Finally, Article XXV addresses the possibility, “in exceptional circumstances not elsewhere provided for” in the agreement, for the signatory countries to “waive an obligation imposed upon a contracting party” (Art XXV §5) by the GATT.

In what follows special attention will be given to Article XIX, because it specifically focuses on situations in which a country suffers from sudden import surges that seriously threaten domestic industries and thus may be exposed to the temptation to break commitments made in the GATT. In order to prevent deviations from the agreement in such situations Article XIX provides the possibility to temporarily suspend obligations. Article XIX §1(a) reads:

If, as a result of unforeseen developments and of the effect of the obligations incurred by a contracting party under this Agreement, including tariff concessions, any product is being imported into the territory of that contracting party in such increased quantities and under such conditions as to cause or threaten serious injury to domestic producers in that territory of like or directly competitive products, the contracting party shall be free, in respect of such product, and to the extent and for such time as may be necessary to prevent or remedy such injury, to suspend the obligation in whole or in part or to withdraw or modify the concession.

In Article XIX §2 the necessity of giving notice of any such action to the contracting parties as far in advance as possible is emphasized. Furthermore the exporters of the products concerned shall be given the opportunity for consultation regarding the action. However, if delay of implementation of the action would be harmful, provisional action without prior consultation is allowed for, granted that consultation is effected immediately after taking such action. In case consultations do not end in agreement, the affected country is nevertheless permitted to take or continue the action, in which case “the affected contracting parties shall then be free, not later than ninety days after such action is taken, to suspend, upon the expiration of thirty days from the day on which written notice of such suspension is received by the contracting parties, the application to the trade of the contracting party taking such action, or ... to the trade of the contracting party requesting such action, of such substantially equivalent concessions or other obligations under this Agreement the suspension of which the contracting parties do not disapprove” (Art XIX §3:a). However, in case a country faces serious injury due to such action, it is “free to suspend, upon the taking of the action and throughout the period of consultation, such concessions or

other obligations as may be necessary to prevent or remedy the injury” (Art XIX §3:b). Hence there is a clear provision for retaliatory action against the application of a safeguard measure under Article XIX.

2.2 The Uruguay Round Agreement on Safeguards

The Tokyo Round (1973-1979) did not produce an agreement on a code of conduct governing the use of safeguard measures pursuant to Article XIX. In the Uruguay Round (1986-1994) however, an agreement on safeguards was achieved.⁵ With special regard to the clarifying and reinforcing the disciplines of Article XIX, the need “to re-establish multilateral control over safeguards and eliminate measures that escape such control” is emphasized. The application of safeguard measures to a product are conditioned on that product being imported “in such increased quantities, absolute or relative to domestic production, and under such conditions as to cause or threaten to cause serious injury to the domestic industry that produces like or directly competitive products” (Art 2). “Serious injury” is defined as “significant overall impairment in the position of a domestic industry” (Art 4:1a), while “threat of serious injury” is defined as clearly imminent serious injury, the determination of which “shall be based on facts and not merely on allegation, conjecture or remote possibility” (Art 4:1b). When determining whether increased imports are causing or threatening to cause serious injury particular emphasis should be put on “the rate and amount of the increase in imports of the product concerned in absolute and relative terms, the share of the domestic market taken by increased imports, changes in the level of sales, production, productivity, capacity utilization, profits and losses, and employment” (Art 4:2a).⁶ Safeguard measures should be applied “only to the extent necessary to prevent or remedy serious injury and to facilitate adjustment” (Art 5:1) and be on a MFN basis, although selective applications are permitted under Article 5:2b.⁷

⁵The Uruguay Round agreements consist of the Agreement Establishing the World Trade Organization and, annexed to it, the agreements on trade in goods (GATT), on trade in services (GATS) and on trade-related aspects of intellectual property rights (TRIPS), the dispute settlement understanding, the trade policy review mechanism and the plurilateral agreements, as well as the schedules of commitments. The multilateral agreements on trade in goods (annex 1A) contain, among others, the Agreement on Safeguards.

⁶Schott (1994) notes that the provisions for serious injury establish a much higher threshold than the “material injury” standard for antidumping and countervailing duties outlined in Article VI.

⁷According to Schott (1994) these selective applications seems to have been aimed at exporters in Asia and Eastern Europe, and in particular China.

Safeguard measures are permitted only for a period not exceeding four years (Art 7:1), except under special circumstances (Art 7:2); the total period of application shall however not exceed eight years (Art 7:3). Whenever it can be expected that a safeguard measure will be applied for more than one year, “the Member applying the measure shall progressively liberalize it at regular intervals during the period of application” (Art 7:4). Moreover, to prevent circumvention of the requirement of liberalization of safeguard measures, it is not permitted to apply a safeguard measure “for a period of time equal to that during which such measure had been previously applied, provided that the period of non-application is at least two years” (Art 7:5).⁸ These clearly defined time limits on the use of a safeguard measure stand in sharp contrast to the previous, somewhat vague constraint for the application of a measure to be “temporary”.

In order to alleviate the effects of exercising a safeguard measure it is required that “a substantially equivalent level of concessions and other obligations to that existing under GATT 1994” is maintained between the measure-taking party and the affected parties, which may be achieved through “any adequate means of trade compensation for the adverse effects of the measure on their trade” (Art 8:1). In case agreement is not reached within 30 days, an affected exporting party has the right to suspend the application of “substantially equivalent concessions or other obligations under GATT 1994” (Art 8:2), but not “for the first three years that a safeguard measure is in effect, provided that the safeguard measure has been taken as a result of an absolute increase in imports and that such a measure conforms to the provisions of this Agreement” (Art 8:3).

Developing countries are granted special treatment. Safeguards are not allowed against a developing country “as long as its share of imports of the product concerned in the importing Member does not exceed 3 per cent, provided that developing country Members with less than 3 per cent import share collectively account for not more than 9 per cent of total imports of the product concerned” (Art 9:1). Moreover, a developing country has the possibility to extend a safeguard measure up to ten years, inclusive of extensions, and can reintroduce a safeguard measure against a product after a period of time equal to only half that

⁸An exception is made if the safeguard measure has a duration of 180 days or less, at least one year has elapsed since the introduction of the initial measure and such a measure has not been applied more than twice during five years preceding the date of introduction of the measure (Art 7:6).

during which the previous was applied, given that two years have passed (Art 9:2).

In order to stem the use of non-orderly quantitative restrictions, “voluntary export restraints, orderly marketing arrangements or any other similar measures on the export or the import side” are explicitly prohibited (Art 11:1b), something that has been considered as one of the greatest achievements of the Uruguay Round.⁹

Schott (1994) notes that before the conclusion of the Uruguay Round Article XIX measures had to be non-discriminatory and affected exporters had the right to claim compensation or to seek authorization for retaliation. Governments thus often preferred other less costly and more flexible safeguard measures, either because they sought to exempt certain countries, or in order to avoid the need for compensation. When assessing the results of the Uruguay Round, Schott (1994) emphasizes the ban on VERs as one of the major achievements, but criticizes the weak incentives for adjustment resulting from the generous durations of safeguard actions, the failure to remove the justification for balance of payments safeguard measures imposed by developing countries and the failure to discipline their use. He notes that the removal of the threat of retaliation for three years, the ban on alternative “gray area” measures such as VERs, the omit of an adjustment requirement for at least four years and the possibility for selective application of safeguards seem to have been aimed to encourage the use of Article XIX. However the reduced risk of retaliation is balanced by the relatively rigid serious injury requirement, the MFN requirement in most cases and the constraint on actions vis-à-vis developing countries. It is therefore likely that safeguard measures will continue to play a minor role, at least compared to antidumping actions.

2.3 Economic Implications of Safeguard Measures

Hoekman and Kostecki (1995) note that safeguard actions distribute income from consumers to import-competing and/or foreign exporting industries. The mere existence of safeguard instruments may reduce competition between foreign exporters and domestic import-competing firms. Scope may also exist for the capture and abuse of such procedures by import-competing interests, further strengthening such effects. Hence the gains from the liberalization negotiated under multilateral trade negotiations or implemented unilaterally are reduced for certain sectors or

⁹See, for example, Schott (1994) and Hoekman and Kostecki (1995).

for the economy as a whole. In so far as the cause of an import-competing industry's problems lies in a shift in comparative advantage, protection is generally an inappropriate policy to bring about adjustment. Whatever the political rationale for safeguard instruments, their mere existence may reduce competitive pressure on domestic import-competing firms, e.g. by raising prices or reducing incentives to innovate. Moreover they are usually economically inefficient, since the costs to consumers are almost always larger than the benefits that accrue to the protected industry. Furthermore, industries can be expected to exploit substitution possibilities across instruments if these exist, making it more difficult for governments to control trade policy.

2.4 The Use of Safeguards

Safeguard measures have always been available under Article XIX of the GATT. However, they have been used infrequently. Some governments have preferred to protect their domestic industries by using bilateral negotiations outside the GATT in order to persuade exporting countries to restrain exports voluntarily or to agree to other means of sharing markets (so-called "gray area" measures). According to Schott (1994), Article XIX was invoked at 151 occasions up to May 1993, almost a third of which after the conclusion of the Tokyo Round in 1979, while for example the number of antidumping cases only between 1985 and 1992 exceeded one thousand.

The dominance of AD and VERs in OECD countries reflects the fact that the requirements for invoking Article XIX have been stringent (Hoekman and Kostecki, 1995). The relatively rare use of Article XIX can be explained by the condition of "serious injury" being too restrictive, the implicit requirement of compensation and the risk of facing retaliation (Schott, 1994). Among developing countries, Article XVIII:b has frequently been used in order to enable quantitative restrictions.

3 A Prisoner's Dilemma Tariff Setting Game

3.1 The Standard Tariff Setting Game

The setting is as follows. There are two countries, each exporting one good to the other, but in all other respects perfectly symmetric. Each country's payoff W is a function of the own tariff t and the foreign tariff t^* , i.e. $W = W(t, t^*)$. There exists a best-response function $t_{BR}(t^*) = \arg \max_t W(t, t^*)$. W increases in t for $t < t_{BR}(t^*)$, while it decreases in t for $t > t_{BR}(t^*)$ (as long as trade takes place). Hence for any given t^*

there is a unique t that maximizes W (as long as trade takes place). W is falling monotonously in t^* (as long as trade takes place).

A simple and frequently used way of modeling trade policy is to regard tariff setting as an infinite repetition of a Prisoner's Dilemma game between two countries that can choose between cooperating or deviating. In fact there are two stages. At the first stage, before the infinitely repeated game begins, the two countries choose a cooperative tariff level t_C from a continuum and agree on how deviations should be punished. At the second stage, the infinitely repeated game is played. Once it starts each country will choose between implementing the agreed-upon cooperative tariff and applying the optimal tariff $t_D = t_{BR}(t_C)$ vis-à-vis the other country. Actually a country can choose its tariff level from a continuum. However, assuming that setting the tariff different from t_C is regarded as a deviation, a country's choice will in fact be binary, i.e. between applying t_C and t_D . Under perfect symmetry, the per-period payoff under cooperation is given by $W_C = W(t_C, t_C)$. A country that decides to break its commitment by applying the optimal tariff vis-à-vis its trading partner gets the payoff $W_D = W(t_D, t_C)$, while its trading partner receives the sucker's payoff $W_S = W(t_C, t_D)$. In the absence of cooperation both countries apply the optimal tariff vis-à-vis each other, i.e. the Nash tariff rate $t_N = t_{BR}(t_{BR})$, and both receive the payoff $W_N = W(t_N, t_N)$. The chosen cooperative tariff level t_C not only directly defines payoff under cooperation (W_C), but also indirectly, via the best-response function, defines payoffs of deviation (W_D) and being deviated-against (W_S).

There exists a unique tariff level $t_C^{opt} < t_N$ that maximizes $W_C = W(t_C, t_C)$. W_C increases at a decreasing rate in t_C for $t_C < t_C^{opt}$, while it decreases at an increasing rate for $t_C > t_C^{opt}$. It follows immediately that there exists a $t' < t_C^{opt}$ such that $W_C = W_N$ for $t_C = t'$. Thus $W_C > W_N$ if and only if $t' < t_C < t_N$.

W_S increases at a decreasing rate in t_C for $t_C < t_N$ and decreases at an increasing rate in t_C for $t_C > t_N$, while W_D decreases in t_C . Letting t_C decrease below t_N thus leads to a monotonous increase in W_D . The increase in W_D is equal to the increase in W_C (i.e. W_D and W_C are tangent) at $t_C = t_N$ and unambiguously stronger for $t_C < t_N$. Thus $W_D - W_C$ increases, and it does so at an increasing rate as t_C decreases.

It follows immediately that $\lim_{t_C \rightarrow t_N} \frac{W_D - W_C}{W_C - W_N} = 0$ and that $\frac{W_D - W_C}{W_C - W_N}$ increases monotonously as t_C decreases below t_N as long as $W_C > W_N$ (obviously $\lim_{t_C \searrow t'} \frac{W_D - W_C}{W_C - W_N} = \infty$). For $\frac{W_D - W_C}{W_D - W_N}$ the following is true:

$\lim_{t_C \rightarrow t_N} \frac{W_D - W_C}{W_D - W_N} = 0$ and $\frac{W_D - W_C}{W_D - W_N}$ increases monotonously as t_C decreases below t_N .

The relevant range of cooperative tariffs to consider are those that are strictly smaller than the Nash tariff rates and larger than the optimal cooperative tariff level, i.e. $t_C^{opt} \leq t_C < t_N$. In this range it is the case that $W_D > W_C > W_N > W_S$. (For $t_C = t_N$ it is obviously the case that $W_D = W_C = W_S = W_N$) The per-period payoff matrix for the symmetric case, given below, is thus of Prisoner's Dilemma type.

	Cooperate	Deviate
Cooperate	W_C, W_C	W_S, W_D
Deviate	W_D, W_S	W_N, W_N

A cooperative outcome in every period can be maintained if countries stick to a grim-trigger strategy, i.e. any deviation by one country will be punished by infinite reversion to the Nash equilibrium. Cooperation is sustainable, if and only if the cost of deviating outweighs the one-period gain from deviating, i.e. if and only if

$$W_D - W_C \leq \frac{\beta}{1 - \beta} [W_C - W_N] \quad (1)$$

where β is the discount factor. The left-hand side represents the short-term (one-period) gain from deviation, while the right-hand side represents the expected long-term loss from deviation. Rearranging terms yields the following equation.

$$\frac{W_D - W_C}{W_C - W_N} \leq \frac{\beta}{1 - \beta} \quad (1')$$

This equation tells us that, in order to sustain cooperation, t_C only can be lowered to the degree that $\frac{W_D - W_C}{W_C - W_N}$ does not exceed the upper bound, which is solely determined by the discount factor and increases monotonously in it. Since $\beta \in (0, 1)$ and thus $\frac{\beta}{1 - \beta} \in (0, \infty)$ it is always possible to find some $t_C < t_N$ that is sustainable. A higher discount factor implies that the upper bound increases, and it is thus possible to sustain a lower t_C . The restriction given by (EQ1') is incorporated into the negotiations concerning the cooperative tariff level, i.e. the payoff under cooperation W_C , and thus imposes an upper limit on cooperation.

Lemma 1 *The optimal cooperative tariff is sustainable for a sufficiently high discount factor.*

Proof. Letting β approach unity, the right-hand side of (1') goes to infinity. Since $\frac{W_D - W_C}{W_C - W_N}$ is finite for $t_C = t_C^{opt}$, it follows immediately that (1') holds for a sufficiently high β . ■

3.2 Introducing Uncertainty in a Repeated Prisoner's Dilemma Game

So far it has been assumed that there is perfect certainty regarding payoffs both in the present and all future periods. If there is certainty about the payoffs in each period, infinite cooperation can easily be sustained, as long as (1) holds. However, unanticipated events may occur. Following Rosendorff and Milner (2001) it will in what follows be assumed that in every period each country is affected by exogenous shocks ε and ε^* respectively. These shocks are assumed to be identically and independently distributed and not observable to the trading partner. Hence each country has perfect knowledge about the shock that it is exposed to, but it knows nothing about the shock that its trading partner is presently experiencing, and neither does it know anything about the shocks that will occur in future periods both at home and abroad.

Following Rosendorff and Milner (2001), a situation in which ex ante both countries are uncertain regarding the precise outcomes in future periods is considered. Both countries know that in each period they will be exposed to a temporary (one period lasting) political economic shock that impacts on the returns from cooperation and deviation during that period, but do not know the magnitude of the shock. It is furthermore assumed that this shock has no influence on future payoffs. What countries negotiate about is therefore the expected value under mutual cooperation, while being aware that the actual outcome will be influenced by a stochastic variable.

There are three channels through which the different payoffs of the game in figure 1 are affected by these variables. First, the payoff functions is affected such that $W^\varepsilon = W(t, t^*; \varepsilon)$. Second, the own best-response function is affected such that $t_D^\varepsilon = t_{BR}^\varepsilon(t_C)$. And third, the best-response function of the trading partner is affected such that $t_D^{*\varepsilon} = t_{BR}^{*\varepsilon}(t_C)$. Thus, while the payoff under cooperation will only be affected through the first channel, i.e. $W_C^\varepsilon = W(t_C, t_C; \varepsilon)$, the payoff under deviation will also be affected through the second channel, i.e. $W_D^\varepsilon = W(t_D^\varepsilon, t_C; \varepsilon)$. The sucker's payoff and the Nash payoff will both be affected through the third channel. Whereas the sucker's payoff is not

affected through the second channel, i.e. $W_{S,\varepsilon^*}^\varepsilon = W(t_C, t_D^{*\varepsilon}; \varepsilon)$, the Nash payoff is affected through all three channels, i.e. $W_{N,\varepsilon^*}^\varepsilon = W(t_D^\varepsilon, t_D^{*\varepsilon}; \varepsilon)$.

For tractability it will be assumed that the best-response functions are unaffected by these exogenous shocks.

$$(A1) \quad t_{BR}^\varepsilon = t_{BR} \quad \text{for any } \varepsilon$$

By making this simplifying assumption the analysis is solely focused upon the effects of a shock through the first channel. Thus the shock that the trading partner is experiencing has no direct effect on the own payoff. It is only the own shock that affects payoffs, i.e. $W = W(t, t^*; \varepsilon)$ and $W^* = W(t, t^*; \varepsilon^*)$. However the shock of the trading partner has an effect on the strategic decision it takes on whether or not to continue cooperation.

Next it will be assumed that the difference in payoffs between deviating and cooperating is independent of the action taken by the trading partner.

$$(A2) \quad W(t_D, t^*; \varepsilon) - W(t_C, t^*; \varepsilon) = \text{constant for all } t^*$$

This implies that $W_D^\varepsilon - W_C^\varepsilon = W(t_D, t_C; \varepsilon) - W(t_C, t_C; \varepsilon) = W(t_D, t^*; \varepsilon) - W(t_C, t^*; \varepsilon)$ for any t^* and in particular that $W_N^\varepsilon - W_S^\varepsilon = W_D^\varepsilon - W_C^\varepsilon$. Thus the short-term gain of deviating is independent of what action the trading partner takes. Making this assumption simplifies things since no consideration has to be taken about what action the trading partner will take. Therefore any decision will solely depend on the domestic shock.

To sustain cooperation the expected outcome under deviation has to be smaller than the expected outcome under cooperation. Let V_C be the continuation value if cooperation is sustained in the present period, and let V_D be the continuation value if deviation occurs. Let p be the probability that the trading partner chooses to cooperate.

$$\begin{aligned}
& C \succ D \\
& \Leftrightarrow p[W_C^\varepsilon + \beta V_C] + (1-p)[W_S^\varepsilon + \beta V_D] \\
& \geq p[W_D^\varepsilon + \beta V_D] + (1-p)[W_N^\varepsilon + \beta V_D] \\
& \Leftrightarrow \beta p[V_C - V_D] \geq p[W_D^\varepsilon - W_C^\varepsilon] + (1-p)[W_N^\varepsilon - W_S^\varepsilon] \\
(A2) \quad & \Leftrightarrow \beta p[V_C - V_D] \geq p[W_D^\varepsilon - W_C^\varepsilon] + (1-p)[W_D^\varepsilon - W_C^\varepsilon] \\
& \Leftrightarrow W_D^\varepsilon - W_C^\varepsilon \leq \beta p[V_C - V_D] \tag{2}
\end{aligned}$$

In the absence of shocks (2) is equivalent to equation (1) being satisfied with an implied probability of cooperating equal to unity ($p = 1$). If

shocks do however occur, the decision to cooperate depends on the size of the one-period gain to be made by deviating. Since present shocks are assumed not to have any impact on expectations concerning future payoffs, the continuation values are based on expected values that are independent of the present situation. Hence the right-hand side of (2) is constant, and it is the size of the left-hand side that determines the choice whether to cooperate or not. What matters is therefore the difference in payoffs between deviating and cooperating under the shock.

Next, it will be assumed that the exogenous variable affects the incentive to deviate in the following way.

$$(A3) \quad W_D^\varepsilon - W_C^\varepsilon = \varepsilon[W_D - W_C]$$

where ε has the density function φ and the cumulative distribution function Φ defined over the interval $[0, \infty]$, and the expected value of ε equals one.¹⁰ By making this assumption regarding the effect of the exogenous variable ε it is taken into account that what matters are not absolute realizations of W_C^ε and W_D^ε , but how the one-period gain from deviating is affected. Thus, a country that gets a lower than expected payoff under cooperation may nevertheless not opt for deviating because the expected payoff under deviation is even lower than expected. And, by the same token, a country that gets a higher than expected payoff from cooperation may nonetheless be inclined to deviate, because the payoff from deviating is even higher than expected. Using (A2) and (A3) it follows immediately that

$$W_N^\varepsilon - W_S^\varepsilon = W_D^\varepsilon - W_C^\varepsilon = \varepsilon[W_D - W_C] = \varepsilon[W_N - W_S]$$

The expected value of the one-period gain from deviating is thus equal to the one in the game without shocks. By defining the distribution function over the interval $[0, \infty]$ it is avoided that some arbitrary assumption regarding the upper bound for the size of the one-period gain from deviating is made. Thus the one-period gain from deviating can potentially take on extremely large values, albeit with probabilities that are infinitesimally small. By letting the shock enter the equation multiplicatively account is taken for the impact of a shock being different for different cooperative tariff levels. Since the one-period gain from deviating is higher the lower the cooperative tariff is set, the same shock will have a stronger impact at a lower cooperative tariff level. If the cooperative tariff is set equal to the Nash tariff ($t_C = t_N$), the payoffs under cooperation, deviation and being deviated against are all equal to

¹⁰One possible distribution could be a log-linear distribution function.

the Nash payoff ($W_C = W_D = W_S = W_N$). In this case the one-period gain from deviating will of course be zero, independent of the size of the shock ε , something that is captured by (A3). Furthermore, for any cooperative tariff level below the Nash tariff, a sufficiently large shock will make deviation worthwhile. Thus, cooperation will in this setting not be sustainable for any $t_C < t_N$ if an unanticipated, sufficiently large shock occurs. The lower the cooperative tariff is set the lower will the threshold value for ε , above which deviation occurs, be. Plugging (A3) into (2) and rearranging terms gives us the following equation.

$$\varepsilon \leq \beta p \frac{V_C - V_D}{W_D - W_C} \equiv \varepsilon'$$

A country will not know what action its trading partner will take but it can estimate the likelihood of cooperation. This estimation has to be consistent in the sense that it is equal to the implied probability of choosing cooperation. By symmetry, which implies that the probability for choosing cooperation is the same for both countries, this probability is therefore given by $p = \text{prob}(\varepsilon \leq \varepsilon') = \Phi(\varepsilon')$. Plugging this into the above equation gives us the following condition for consistent solutions.

$$\beta \Phi(\varepsilon') \frac{V_C - V_D}{W_D - W_C} = \varepsilon' \quad (3)$$

Consistency in beliefs require (3) to hold. Hence given a certain t_C and thus a certain value for $W_D - W_C$, arriving at a threshold value ε' requires taking into account the consistency of this value, i.e. that it is consistent with its implied value given by (3).

Lemma 2 *There exists at least one solution to equation (3).*

Proof. Since $\Phi(0) = 0$, it follows immediately that $\varepsilon' = 0$ solves (3) for any $t_C < t_N$. ■

Never to cooperate, independent of the shock a country is experiencing, is therefore always a consistent strategy. Believing that the trading partner will never opt for cooperation, the best response is never to choose cooperation either. There may or may not be further solutions, depending on the discount factor, the distribution function and the expected one-period gain from deviating (i.e. the cooperative tariff). In order to assess the impacts of these factors it is necessary to determine the continuation values. Given that grim-trigger strategies are applied, i.e. deviation is followed by infinite reversion to the Nash equilibrium,

the continuation values are given by (for the derivation of V_C , see the appendix):

$$V_D = \frac{W_N}{1 - \beta} \quad (4)$$

$$V_C = V_D + \frac{1}{1 - \beta\Phi^2(\varepsilon')} \{ [W_D - W_N]\Phi(\varepsilon') - [W_D - W_C] \int_0^{\varepsilon'} \varepsilon d\Phi \} \quad (5)$$

A lowering of the cooperative tariff level affects the continuation value under cooperation in two ways. First it directly affects the terms $W_D - W_N$ and $W_D - W_C$. Second it influences the threshold value ε' and thus the probability of choosing cooperation. The first effect is unambiguously positive for $t_C \in [t_C^{opt}, t_N]$ as some rearrangements of (5) demonstrate.

$$\begin{aligned} V_C - V_D &= \frac{1}{1 - \beta\Phi^2(\varepsilon')} \{ [W_D - W_N]\Phi(\varepsilon') - [W_D - W_C] \int_0^{\varepsilon'} \varepsilon d\Phi \} \\ &= \frac{1}{1 - \beta\Phi^2(\varepsilon')} \{ [W_C - W_N]\Phi(\varepsilon') + [\Phi(\varepsilon') - \int_0^{\varepsilon'} \varepsilon d\Phi][W_D - W_C] \} \\ &= \frac{\Phi(\varepsilon')}{1 - \beta\Phi^2(\varepsilon')} \{ [W_C - W_N] + [1 - E(\varepsilon|\varepsilon \leq \varepsilon')][W_D - W_C] \} \quad (6) \end{aligned}$$

For $t_C \in [t_C^{opt}, t_N]$ a decrease in t_C leads to an increase in $W_C - W_N$ and in $W_D - W_C$ and thus to an increase in $V_C - V_D$. The second effect is somewhat harder to assess. Decreasing t_C increases $W_D - W_C$ which decreases the right-hand side of (3). If however $V_C - V_D$ also increases this effect will be balanced. The total effect is likely to be a decrease in ε' . This in turn will probably lead to a fall in $V_C - V_D$, since the effect on $\frac{\Phi(\varepsilon')}{1 - \beta\Phi^2(\varepsilon')}$ will be stronger than on $E(\varepsilon|\varepsilon \leq \varepsilon')$. Hence there is a tradeoff: the lower the cooperative tariff is set, the less likely is the continuation of cooperation. In order to establish the optimal cooperative tariff level in the presence of uncertainty it is necessary to find out the exact interrelation between the cooperative tariff level and the probability to cooperate. What matters here is the concept of consistency. The belief about the probability of the trading partner opting for cooperation has to be consistent with the own threshold value for the exogenous variable below which cooperation is chosen. This consistency is expressed in equation (3). In lemma 1 it was established that there exists at least

one solution to (3); never to choose cooperation is a consistent solution. In order to find out whether there are any further consistent solutions (5) has to be plugged into (3).

$$\begin{aligned}\varepsilon' &= \frac{\beta\Phi(\varepsilon')}{1 - \beta\Phi^2(\varepsilon')} \frac{1}{W_D - W_C} \{ [W_D - W_N]\Phi(\varepsilon') - [W_D - W_C] \int_0^{\varepsilon'} \varepsilon d\Phi \} \\ &= \frac{\beta\Phi(\varepsilon')}{1 - \beta\Phi^2(\varepsilon')} \left\{ \frac{W_D - W_N}{W_D - W_C} \Phi(\varepsilon') - \int_0^{\varepsilon'} \varepsilon d\Phi \right\} \equiv f(\varepsilon')\end{aligned}\quad (7)$$

The left-hand side of (7) takes on values in the interval $[0, \infty]$, while the right-hand takes on values in a positive, bounded interval.

Proposition 3 *There exists at least one stable solution ($\varepsilon' = 0$) to equation (7).*

Proof. (i) Lemma 2 $\Rightarrow \varepsilon' = 0$ is a solution to (7)

(ii) The derivative of $f(\varepsilon')$ is given by

$$\begin{aligned}f'(\varepsilon') &= \frac{\beta\Phi(\varepsilon')}{1 - \beta\Phi^2(\varepsilon')} \left\{ \frac{W_D - W_N}{W_D - W_C} \varphi(\varepsilon') - \varphi(\varepsilon')\varepsilon' \right\} \\ &\quad + \left\{ \frac{2\beta^2\Phi^2(\varepsilon')\varphi(\varepsilon')}{[1 - \beta\Phi^2(\varepsilon')]^2} + \frac{\beta\varphi(\varepsilon')}{1 - \beta\Phi^2(\varepsilon')} \right\} \left\{ \frac{W_D - W_N}{W_D - W_C} \Phi(\varepsilon') - \int_0^{\varepsilon'} \varepsilon d\Phi \right\} \\ &= \frac{\beta\Phi(\varepsilon')\varphi(\varepsilon')}{1 - \beta\Phi^2(\varepsilon')} \left\{ \frac{W_D - W_N}{W_D - W_C} - \varepsilon' \right\} \\ &\quad + \frac{1}{[1 - \beta\Phi^2(\varepsilon')]^2} \left\{ \beta^2\Phi^2(\varepsilon')\varphi(\varepsilon') + \beta\varphi(\varepsilon') \right\} \left\{ \frac{W_D - W_N}{W_D - W_C} \Phi(\varepsilon') - \int_0^{\varepsilon'} \varepsilon d\Phi \right\} \\ &= \frac{\beta\Phi(\varepsilon')\varphi(\varepsilon')}{[1 - \beta\Phi^2(\varepsilon')]^2} \left\{ \frac{W_D - W_N}{W_D - W_C} - \varepsilon' \right\} \{1 - \beta\Phi^2(\varepsilon')\} \\ &\quad + \frac{\beta\Phi(\varepsilon')\varphi(\varepsilon')}{[1 - \beta\Phi^2(\varepsilon')]^2} \{1 + \beta\Phi^2(\varepsilon')\} \left\{ \frac{W_D - W_N}{W_D - W_C} - E(\varepsilon|\varepsilon \leq \varepsilon') \right\} \\ &= \frac{\beta\Phi(\varepsilon')\varphi(\varepsilon')}{[1 - \beta\Phi^2(\varepsilon')]^2} \left\{ 2\frac{W_D - W_N}{W_D - W_C} - [1 - \beta\Phi^2(\varepsilon')]\varepsilon' - [1 + \beta\Phi^2(\varepsilon')]E(\varepsilon|\varepsilon \leq \varepsilon') \right\}\end{aligned}$$

$\Rightarrow f'(0) = 0$ and $f'(\varepsilon') \geq 0$ and $f(\varepsilon') < \varepsilon'$ close to $\varepsilon' = 0$ for any $t_C < t_N$

$\Rightarrow \varepsilon' = 0$ is a stable solution ■

Letting ε' go to infinity, f converges to $\frac{\beta}{1-\beta} \frac{W_D - W_N}{W_D - W_C}$. There may or may not exist further solutions, depending on the discount factor and on the level of the cooperative tariff.

By increasing the discount factor, $f(\varepsilon')$ increases for any given ε' . Since for $f(\varepsilon')$ close to unity letting β approach unity will make the term $\frac{\beta \Phi(\varepsilon')}{1-\beta \Phi^2(\varepsilon')}$ very large, it is the case that for a sufficiently high discount factor there will be values of ε' , for which $f(\varepsilon') \geq \varepsilon'$. Similarly, by decreasing the discount factor, $f(\varepsilon')$ decreases for any given ε' , and eventually $f(\varepsilon') < \varepsilon'$ for all $\varepsilon' > 0$. Thus there exists a threshold value of β , below which there exist no further solutions for equation (7).

The reasoning is similar for the chosen cooperative tariff level. The term $\frac{W_D - W_N}{W_D - W_C}$ goes to infinity as $t_C \rightarrow t_N$. By letting t_C decrease, this term falls and eventually goes to zero as $t_C \rightarrow -\infty$. Thus, by decreasing t_C , $f(\varepsilon')$ decreases for any given ε' . If t_C is sufficiently low, i.e. $\frac{W_D - W_N}{W_D - W_C}$ is sufficiently low, $f(\varepsilon') < \varepsilon'$ for all $\varepsilon' > 0$. In this case there will exist no further solution to equation (7). However for t_C sufficiently close to t_N there will at least be an interval of values for ε' , for which $f(\varepsilon') \geq \varepsilon'$.

Proposition 4 *Among the solutions to (7), the highest will be chosen since it is both stable and renders the highest continuation value.*

Proof. See appendix. ■

To conclude, given a certain discount factor, cooperation can be sustained for t_C sufficiently close to t_N as long as the realizations of the stochastic variable are smaller than a threshold value. Cooperation will however break down as soon as this threshold value is exceeded. When t_C is too small, cooperation will break down instantly, since in this case each country will be sure that its trading partner will deviate and thus decide to deviate as well.

4 The Rationale for Introducing an Escape Clause

If both countries know in advance that they might be exposed to the temptation to deviate from cooperation, they might ex ante agree on a mechanism for dealing with situations, in which shocks that induce changes in the payoff occur.

In what follows a modification of the approach taken by Rosendorff and Milner (2001) will be applied. Using the setting described above, an agreement on cooperation could then include the possibility to make use of an escape clause, in case a country faces a severe shock, thereby allowing this country to offset the effects of the shock by deviating temporarily, i.e. during the period during which the shock occurs, and then return to the cooperative regime. It is however necessary that the country that uses the escape clause and thus temporarily deviates incurs some cost, either by imposing the cost upon itself or by being exposed to some action by its trading partner. What matters is that by exercising the escape clause and incurring a cost for doing so a shock-affected country signals its willingness to return to the cooperative regime by making a voluntary concession. Hence, instead of deviating, receiving the one-period payoff W_D and thereby make further cooperation impossible, the country deviates and incurs a cost $F = F(\varepsilon) > 0$, thereby receiving the payoff $W_D - F(\varepsilon)$.¹¹ The other country therefore receives the sucker's payoff W_S plus a compensation $G = G(F, \varepsilon)$ that is non-negatively related to the cost associated with exercising the escape clause, i.e. $\frac{\partial G}{\partial F} \geq 0$.¹² The payoff matrix now looks as follows.¹³

	C^*	EC^*	D^*
C	$W_C^\varepsilon,$ $W_C^{\varepsilon^*}$	$W_S^\varepsilon + G(F, \varepsilon^*),$ $W_D^{\varepsilon^*} - F(\varepsilon^*)$	$W_S^\varepsilon,$ $W_D^{\varepsilon^*}$
EC	$W_D^\varepsilon - F(\varepsilon),$ $W_S^{\varepsilon^*} + G(F, \varepsilon)$	$W_N^\varepsilon + G(F, \varepsilon^*) - F(\varepsilon),$ $W_N^{\varepsilon^*} + G(F, \varepsilon) - F(\varepsilon^*)$	$W_N^\varepsilon - F(\varepsilon),$ $W_N^{\varepsilon^*} + G(F, \varepsilon)$
D	$W_D^\varepsilon,$ $W_S^{\varepsilon^*}$	$W_N^\varepsilon + G(F, \varepsilon^*),$ $W_N^{\varepsilon^*} - F(\varepsilon^*)$	$W_N^\varepsilon,$ $W_N^{\varepsilon^*}$

The sequence of events during a period is as follows:

1. At the beginning of a period, both countries experience independent shocks that are unobservable to the trading partner.
2. Both countries determine, independently, what policy they will apply, the options being cooperation, use of the escape clause, and deviation.

¹¹Note that for simplicity it is assumed that the cost for deviating is independent of the degree of deviation. Hence optimal deviation will always be chosen.

¹²In Rosendorff and Milner (2001) there is just a cost for the country that uses the escape clause. The trading partner receives no compensation in their model.

¹³Note that for $F > 0$ the only Nash equilibrium of the one-period game is (D,D*).

3. Both countries implement their policies, and the period begins.
4. At the end of the period the implemented policies are verified. Any deviation by one country is regarded as a breach of the agreement and therefore lead to the breakdown of the cooperative regime.

For tractability the following simplifying assumptions will be made.

$$(A4) \quad G(F, \varepsilon) = F(\varepsilon)$$

Assumption (A4), implying that the cost that the country exercising the escape clause incurs translates into a benefit of exactly the same size for its trading partner is strong, especially if compensation were in the form of a decrease in some other tariff or an increase in the other country's tariff. It is however not unrealistic. It is valid if compensation is purely monetary, i.e. the country exercising the escape clause is required to transfer money to its trading partner, a case that will be examined more closely in the next section.

First the threshold value, below which cooperation is preferred to deviation, is derived. Let p_C and p_{EC} be the probabilities that the trading partner opts for cooperation and exercising the escape clause respectively, and let V'_C be the continuation value of the game if no deviation occurs in the present period. As before $V_D = \frac{W_N}{1-\beta}$ is the continuation value if deviation occurs and is followed by infinite Nash reversion.

$$C \succ D$$

$$\begin{aligned} & \Leftrightarrow p_C\{W_C^\varepsilon + \beta V'_C\} + p_{EC}\{W_S^\varepsilon + \beta V'_C\} + (1 - p_C - p_{EC})\{W_S^\varepsilon + \beta V_D\} \\ & \geq p_C\{W_D^\varepsilon + \beta V_D\} + p_{EC}\{W_N^\varepsilon + \beta V_D\} + (1 - p_C - p_{EC})\{W_N^\varepsilon + \beta V_D\} \\ & \Leftrightarrow p_C[W_C^\varepsilon - W_D^\varepsilon] + (1 - p_C)[W_S^\varepsilon - W_N^\varepsilon] + \beta(p_C + p_{EC})[V'_C - V_D] \geq 0 \\ (A2) \quad & \Leftrightarrow W_C^\varepsilon - W_D^\varepsilon + \beta(p_C + p_{EC})[V'_C - V_D] \geq 0 \\ (A3) \quad & \Leftrightarrow \varepsilon[W_C - W_D] + \beta(p_C + p_{EC})[V'_C - V_D] \geq 0 \\ & \Leftrightarrow \varepsilon \leq \beta(p_C + p_{EC}) \frac{V'_C - V_D}{W_D - W_C} \end{aligned}$$

Next a condition for the compensation cost ensuring that deviation renders a lower outcome than the use of the escape clause is derived.

$EC \succ D$

$$\begin{aligned}
&\Leftrightarrow p_C\{W_D^\varepsilon - F(\varepsilon) + \beta V_C'\} + p_{EC}\{W_N^\varepsilon - F(\varepsilon) + F(\varepsilon^*) + \beta V_C'\} \\
&\quad + (1 - p_C - p_{EC})\{W_N^\varepsilon - F(\varepsilon) + \beta V_C'\} \\
&\geq p_C\{W_D^\varepsilon + \beta V_D\} + p_{EC}\{W_N^\varepsilon + F(\varepsilon^*) + \beta V_D\} \\
&\quad + (1 - p_C - p_{EC})\{W_N^\varepsilon + \beta V_D\} \\
&\Leftrightarrow F(\varepsilon) \leq (p_C + p_{EC})\beta[V_C' - V_D]
\end{aligned}$$

If this condition is satisfied for $\varepsilon > \varepsilon^\circ$, deviation will never be an option, and hence $p_{EC} = 1 - p_C$. In this case the above conditions can be expressed as follows.

$$C \succ D \Leftrightarrow \varepsilon \leq \beta \frac{V_C' - V_D}{W_D - W_C} \equiv \varepsilon^\circ \quad (8)$$

$$EC \succ D \Leftrightarrow F(\varepsilon) \leq \beta[V_C' - V_D] = \varepsilon^\circ[W_D - W_C] \equiv F^\circ \quad (9)$$

Equation (9) tells us that, in order to avoid deviation, the cost for using the escape clause cannot be too high for shocks above the threshold value ε° . given by (8). More specifically, this cost cannot be larger than the discounted future gain from sustaining cooperation vis-à-vis infinite Nash reversion. The threshold value for the shock, above which this restriction has to hold, is given by the ratio between the discounted future gain from sustaining cooperation vis-à-vis infinite Nash reversion and the expected one-period gain from deviating. Thus, to avoid deviations, the escape clause should be used whenever the one-period gain from deviating exceeds the discounted future gain from sustaining cooperation vis-à-vis infinite Nash reversion.

Conclusion 5 *In order to avoid deviations in the presence of unanticipated, temporary shocks, an agreement could include an escape clause, allowing a shock-affected country to temporarily deviate while incurring a cost. For shocks that exceed a threshold value this cost can however not be larger than the discounted value of the expected future gain from sustaining cooperation.*

5 The Optimal Design of an Agreement with an Escape Clause and Monetary Compensation

5.1 The Incentive Problem

Implementing an escape clause cost scheme $F(\varepsilon)$, exactly prescribing the cost associated with exercising an escape clause at a certain shock level, will depend on the observability of shocks. If both ε and the implementation of $F(\varepsilon)$ are perfectly observable, any transfer scheme $F(\varepsilon)$ satisfying (9) for $\varepsilon > \varepsilon^\circ$ will ensure that deviation never occurs. Here it is however assumed that information about the shock is asymmetric, which leads to incentive problems. Depending on what compensation mechanism is chosen, there may be strong incentives to under- or over-estimate the size of a shock, both for the shock-affected country and the escape-clause-exposed country.

There is a great degree of freedom in deciding the shape of the compensation scheme $F(\varepsilon)$, provided the above conditions are met. As can be expected, the range for ε , in which cooperation will be chosen increases in F , the cost associated with exercising the escape clause.¹⁴ A country will prefer F to be low, in order to be able to increase the range within which the escape clause can be exercised and not to have to pay so much compensation in case it faces strong incentives to deviate. However, in case its trading partner faces a shock, it will prefer F to be high, in order to decrease the range within which the escape clause can be exercised and to get a high compensation. Ex ante, when a country knows it can be in either position in future, these two effects on payoff will have to be weighed against each other.

If compensation is monetary and independent of the size of a shock, i.e. $F(\varepsilon) = F$ for any ε , the possibility to cheat is avoided. Were compensation not fixed and depending on the size of a shock, the shock-affected country would overstate ε if $\frac{\partial F}{\partial \varepsilon} < 0$ and understate ε if $\frac{\partial F}{\partial \varepsilon} > 0$. Thus asymmetric information requires monetary compensation to be constant in order to avoid incentive compatibility problems.

¹⁴See also Rosendorff and Milner (2001).

5.2 The Optimal Compensation Cost

Next, the threshold value, below which cooperation is chosen and above which the escape clause is exercised, is derived.

$$\begin{aligned}
& C \succ EC \\
& \Leftrightarrow p_C \{W_C^\varepsilon - [W_D^\varepsilon - F]\} + (1 - p_C) \{W_S^\varepsilon - [W_N^\varepsilon - F]\} \geq 0 \\
& \Leftrightarrow p_C [W_C^\varepsilon - W_D^\varepsilon] + (1 - p_C) [W_S^\varepsilon - W_N^\varepsilon] + F \geq 0 \\
\text{(A2)} \quad & \Leftrightarrow W_C^\varepsilon - W_D^\varepsilon + F \geq 0 \\
\text{(A3)} \quad & \Leftrightarrow F \geq \varepsilon [W_D - W_C] \\
& \Leftrightarrow \varepsilon \leq \frac{F}{W_D - W_C} \equiv \varepsilon_{EC} \tag{10}
\end{aligned}$$

It follows immediately from (9) that $\varepsilon_{EC} \leq \varepsilon^\circ$ holds. Since a higher cost F of exercising the escape clause will constrain its use, the threshold value increases in F . The expected per-period payoff $W_E = (1 - \beta)V'_C$ is given by (for derivation see appendix):

$$W_E = W_N + [W_D - W_N] \Phi(\varepsilon_{EC}) - [W_D - W_C] \int_0^{\varepsilon_{EC}} \varepsilon d\Phi \tag{11}$$

The fact that the compensation cost F has no direct effect on the expected per-period outcome is hardly surprising, because ex ante the probability of using the escape clause and thus incurring the cost for doing so is equal to the probability of being exposed to the use of the escape clause and thus receiving compensation of the same size. The compensation cost F nevertheless has an indirect effect on the expected per-period outcome via its influence on the threshold value ε_{EC} . It is therefore important to find out how F should be set in order to achieve the highest possible expected per-period payoff.

Proposition 6 *The optimal threshold value, above which the escape clause is used, is given by*

$$\varepsilon_{EC}^{opt} = \frac{W_D - W_N}{W_D - W_C} \tag{12}$$

while the optimal compensation cost is given by

$$F^{opt} = \varepsilon_{EC}^{opt} [W_D - W_C] = W_D - W_N \tag{13}$$

Proof. Given a certain cooperative tariff t_C the optimality condition for equation (11) is given by

$$\begin{aligned}\frac{\partial W_E}{\partial \varepsilon_{EC}} &= [W_D - W_N]\varphi(\varepsilon_{EC}) - \varepsilon_{EC}\varphi'(\varepsilon_{EC})[W_D - W_C] \\ &= \varphi(\varepsilon_{EC})\{[W_D - W_N] - \varepsilon_{EC}[W_D - W_C]\} = 0 \\ &\Leftrightarrow \varepsilon_{EC} = 0 \text{ or } \varepsilon_{EC} \rightarrow \infty \text{ or } \varepsilon_{EC} = \frac{W_D - W_N}{W_D - W_C}\end{aligned}$$

The second derivative is given by

$$\frac{\partial^2 W_E}{\varepsilon_{EC}^2} = \{[W_D - W_N] - \varepsilon_{EC}[W_D - W_C]\}\varphi'(\varepsilon_{EC}) - [W_D - W_C]\varphi''(\varepsilon_{EC})$$

For the ε_{EC} that satisfy $\frac{\partial W_E}{\partial \varepsilon_{EC}} = 0$ the second derivatives thus have the following values:

$$\begin{aligned}\lim_{\varepsilon_{EC} \rightarrow 0} \frac{\partial^2 W_E}{\varepsilon_{EC}^2} &= [W_D - W_N]\frac{\varphi'(\varepsilon_{EC})}{\varepsilon_{EC}} \geq 0 \\ \frac{\partial^2 W_E}{\varepsilon_{EC}^2} \Big|_{\varepsilon_{EC} = \frac{W_D - W_N}{W_D - W_C}} &= -[W_D - W_C]\varphi''\left(\frac{W_D - W_N}{W_D - W_C}\right) < 0 \\ \lim_{\varepsilon_{EC} \rightarrow \infty} \frac{\partial^2 W_E}{\varepsilon_{EC}^2} &= 0\end{aligned}$$

Thus, for a certain cooperative tariff, W_E has a unique maximum for $\varepsilon_{EC}^{opt} = \frac{W_D - W_N}{W_D - W_C}$, which, using equation (10), translates into $F^{opt} = W_D - W_N$ as the optimal value for the compensation cost. ■

Since $\lim_{t_C \nearrow t_N} \frac{W_D - W_N}{W_D - W_C} = \infty$ and $\frac{W_D - W_N}{W_D - W_C}$ falls as t_C decreases below t_N , it follows that ε_{EC}^{opt} goes to infinity as t_C approaches t_N and falls as t_C decreases. As t_C decreases, ε_{EC}^{opt} goes to infinity. For $t' < t_C < t_N$, i.e. when $W_C > W_N$, its value is strictly larger than unity, which is the expected value for shocks.¹⁵ Hence, a decrease in the cooperative tariff is associated with a rapid fall in the optimal threshold value above which the escape clause is used. Thus lower cooperative tariffs will lead to an increase in the use of the escape clause when the optimal compensation cost can be applied.

Since $W_D - W_N = 0$ for $t_C = t_N$ and $W_D - W_N$ increases as t_C falls, it follows that F^{opt} increases as t_C falls. Thus lower cooperative tariffs will lead to a higher optimal compensation cost for using the escape clause. And, as noted above, this rising cost will nevertheless be accompanied by an increasing use of the escape clause.

¹⁵Note that ε_{EC}^{opt} is independent of the discount factor.

Conclusion 7 *Letting the cooperative tariff fall below the Nash tariff level leads to an increase in the optimal compensation cost for exercising the escape clause and, although using the escape clause becomes costlier, to an increase in its use.*

As is demonstrated below the optimal compensation cost F^{opt} is exactly equal to the difference between the expected outcome of choosing to use the escape clause (without receiving any compensation if the trading partner uses the escape clause) and the expected outcome of choosing to cooperate (without receiving any compensation if the trading partner uses the escape clause).

$$\begin{aligned}
\text{Expected loss from use of EC by TP} &= \int_0^{\varepsilon_{EC}^{opt}} [W_C^\varepsilon - W_S^\varepsilon] d\Phi + \int_{\varepsilon_{EC}^{opt}}^{\infty} [W_D^\varepsilon - W_N^\varepsilon] d\Phi \\
\text{(A2)} \quad &= \int_0^{\varepsilon_{EC}^{opt}} [W_D^\varepsilon - W_N^\varepsilon] d\Phi + \int_{\varepsilon_{EC}^{opt}}^{\infty} [W_D^\varepsilon - W_N^\varepsilon] d\Phi \\
&= \int_0^{\infty} [W_D^\varepsilon - W_N^\varepsilon] d\Phi \\
&= W_D - W_N = F^{opt}
\end{aligned}$$

The first term equals the difference between the expected payoff of mutual cooperation and the expected payoff of being exposed to the use of the escape clause by the trading partner without receiving compensation, given that the domestic shock is below the threshold level (in which case cooperation is chosen). The second term equals the difference between the expected payoff of unilaterally exercising the escape clause and the expected payoff of mutual use of the escape clause without receiving any compensation from the trading partner, given that the domestic shock is above the threshold level (in which case using the escape clause is chosen). The sum of these two terms therefore equals the difference between the expected outcome if the trading partner cooperates and the expected outcome if the trading partner exercises the escape clause without paying any compensation. Thus the optimal compensation cost is such that it equals the expected loss from being exposed to the use of the escape clause by the trading partner, i.e. on average exposure to the use of the escape clause is fully compensated.

Conclusion 8 *The optimal compensation cost for using the escape clause is such that it equals the expected loss that the trading partner incurs by its use, i.e. on average the trading partner is fully compensated.*

Ex post it may however be the case that a country that has been exposed to the use of the escape clause by its trading partner is adversely affected. If it cooperates this is the case if $W_C^\varepsilon - W_S^\varepsilon - F > 0$ or $F < W_C^\varepsilon - W_S^\varepsilon$, and if it deviates this is the case if $W_D^\varepsilon - F - W_N^\varepsilon > 0$ or $F < W_D^\varepsilon - W_N^\varepsilon = W_C^\varepsilon - W_S^\varepsilon$ (by (A2)). Hence, when a country is exposed to the use of the escape clause by its trading partner will it be undercompensated if $W_D^\varepsilon - W_N^\varepsilon = W_C^\varepsilon - W_S^\varepsilon > F$ and overcompensated if $W_D^\varepsilon - W_N^\varepsilon = W_C^\varepsilon - W_S^\varepsilon < F$; on average however will the compensation be such that it exactly offsets the effects of the escape clause. Nevertheless ex post it is possible that invoking the escape clause turns out not to be socially optimal. Hence conflicts about the use of the escape clause may arise after its implementation.

5.3 The Optimal Compensation Cost Scheme

Given that there is an upper bound for the compensation cost to ensure that the use of the escape clause is preferred to deviation, it is of course not always possible to implement the optimal compensation cost. If $F^{opt} \leq F^\circ$, then the optimal compensation cost can be implemented and the highest possible per-period outcome can be achieved. If however $F^{opt} > F^\circ$, it is not possible to apply the optimal cost. Since W_E is strictly increasing in ε_{EC} for $\varepsilon_{EC} \in (0, \varepsilon_{EC}^{opt})$ and hence in F for $F \in (0, F^{opt})$, it follows immediately that F should be set as close to F^{opt} as possible, i.e. $F = F^\circ$.

Conclusion 9 *The cost scheme should be such that the escape-clause-exposed country gets compensation equal to the expected loss that the trading partner incurs, i.e. $F = W_D - W_N = W_C - W_S$, or gets compensation to the largest possible degree, i.e. $F = F^\circ$. Hence $F = \min(W_D - W_N, F^\circ)$.*

Whether or not it is possible to impose the optimal cost depends on the cooperative tariff and on the discount factor. The threshold value of the discount factor, above which optimality is achievable, is given below (see the appendix for derivation). Optimality is possible whenever the optimal compensation cost is smaller than its upper bound.

$$F^{opt} \leq F^\circ \Leftrightarrow \beta \geq \frac{1}{1 + \Phi(\varepsilon_{EC}^{opt}) \left[1 - \frac{E(\varepsilon | \varepsilon \leq \varepsilon_{EC}^{opt})}{\varepsilon_{EC}^{opt}} \right]} \equiv \beta^\circ \quad (14)$$

It is easily verified that $\frac{1}{2} < \beta^\circ < 1$ for $t_C < t_N$. Letting ε_{EC}^{opt} approach infinity, β° converges to $\frac{1}{2}$. As ε_{EC}^{opt} converges to zero, β° converges to unity.

Conclusion 10 *If the discount factor is sufficiently high, i.e. $\beta \geq \beta^\circ$, the optimal cost for exercising the escape clause can be applied. When the discount factor is not sufficiently high, i.e. $\beta < \beta^\circ$, the highest possible cost, F° , should be applied.*

Next, the interrelation of β° and the cooperative tariff is investigated. As t_C approaches t_N , ε_{EC}^{opt} goes to infinity and β° converges to $\frac{1}{2}$. As t_C approaches $-\infty$, ε_{EC}^{opt} converges to zero and β° converges to unity. Below it is demonstrated that β° decreases unambiguously in ε_{EC}^{opt} .

$$\begin{aligned} \frac{\partial \beta^\circ}{\partial \varepsilon_{EC}^{opt}} &= - \frac{\varphi(\varepsilon_{EC}^{opt}) - \varphi(\varepsilon_{EC}^{opt}) + \frac{\int_0^{\varepsilon_{EC}^{opt}} \varepsilon d\Phi}{\varepsilon_{EC}^{opt} 2}}{\left\{1 + \Phi(\varepsilon_{EC}^{opt}) - \frac{\int_0^{\varepsilon_{EC}^{opt}} \varepsilon d\Phi}{\varepsilon_{EC}^{opt}}\right\}^2} \\ &= - \frac{\int_0^{\varepsilon_{EC}^{opt}} \varepsilon d\Phi}{\left\{\varepsilon_{EC}^{opt} [1 + \Phi(\varepsilon_{EC}^{opt})] - \int_0^{\varepsilon_{EC}^{opt}} \varepsilon d\Phi\right\}^2} < 0 \quad (\text{since } \varepsilon_{EC}^{opt} > 0) \end{aligned}$$

Since a fall in t_C leads to a decrease in $\frac{W_D - W_N}{W_D - W_C}$, it follows immediately that as t_C falls ε_{EC}^{opt} decreases and thus β° increases.

Conclusion 11 *A lower cooperative tariff increases the threshold value for the discount factor, above which the optimal cost for using the escape clause can be applied.*

To summarize the previous findings, given a certain discount factor and a certain cooperative tariff, the best possible compensation cost and the resulting threshold value, above which the escape clause will be exercised, are given by $F_{EC} = \min(F^{opt}, F^\circ)$ and $\varepsilon_{EC} = \min(\varepsilon_{EC}^{opt}, \varepsilon^\circ)$.

It follows immediately from results 4 and 5 that for a given discount factor the following cases can be distinguished. If $\beta^\circ \leq \frac{1}{2}$ the optimal compensation cost cannot be implemented for any $t_C < t_N$; the highest possible compensation cost F° will be implemented. If $\beta^\circ > \frac{1}{2}$, equation

(14) tells us that the optimal compensation cost is implementable as long as

$$\beta \geq \frac{1}{1 + \Phi(\varepsilon_{EC}^{opt}) - \frac{\int_0^{\varepsilon_{EC}^{opt}} \varepsilon d\Phi}{\varepsilon_{EC}^{opt}}} \equiv \beta^\circ(\varepsilon_{EC}^{opt})$$

Since $\beta^\circ = \frac{1}{2}$ for $t_C = t_N$ and β° unambiguously decreases in ε_{EC}^{opt} and hence increases as t_C falls, it follows that for a given $\beta > \frac{1}{2}$ it will be the case that $\beta^\circ \leq \beta$ and thus $F^{opt} \leq F^\circ$ for t_C sufficiently close to t_N ; for lower values of t_C , i.e. when $\beta^\circ(\varepsilon_{EC}^{opt}) > \beta$, it will however be the case that $F^{opt} > F^\circ$ and the optimal compensation cost will not be implementable.

Conclusion 12 *If $\beta > \frac{1}{2}$ then for cooperative tariffs sufficiently close to the Nash tariff the optimal compensation cost is implementable. For lower cooperative tariffs this will however not be the case. If $\beta \leq \frac{1}{2}$ then the optimal compensation cost cannot be applied for any cooperative tariff.*

When $\beta > \frac{1}{2}$, the threshold level for the cooperative tariff, below which $F^{opt} \leq F^\circ$ and above which $F^{opt} > F^\circ$, is implicitly defined by (see the appendix for the derivation):

$$\Phi(\varepsilon_{EC}^{opt}) \left[1 - \frac{E(\varepsilon | \varepsilon \leq \varepsilon_{EC}^{opt})}{\varepsilon_{EC}^{opt}} \right] = \frac{1 - \beta}{\beta} \quad (15)$$

Since $\frac{E(\varepsilon | \varepsilon \leq \varepsilon_{EC}^{opt})}{\varepsilon_{EC}^{opt}} = 1$ for $\varepsilon_{EC}^{opt} = 0$, monotonously decreases in ε_{EC}^{opt} and converges to zero as ε_{EC}^{opt} goes to infinity, it follows that the left-hand side of (15) is monotonously increasing in ε_{EC}^{opt} . Because $\beta > \frac{1}{2}$, the right-hand side of (15) is strictly smaller than one. Hence there exists exactly one value for ε_{EC}^{opt} that solves (15). This value implicitly defines the threshold level t_C , implicitly defined by (12), below which the optimal cost cannot be implemented and above which that is possible.

It is easy to see that as the discount factor increases, the right-hand side of (15) decreases and the solution to (15) decreases as well. Hence, as β increases, the threshold value for t_C must decrease.

5.4 Maximizing Expected Per-period Payoffs

Obviously finding the cooperative tariff that maximizes expected per-period payoffs depends crucially on whether or not the optimal compensation cost can be implemented. From Result 7 we know that if $\beta \leq \frac{1}{2}$ the optimal compensation cost cannot be implemented. If $\beta > \frac{1}{2}$ it is however implementable for sufficiently high t_C . In order to find the optimal cooperative tariff these two cases have to be dealt with independently.

5.4.1 When the Optimal Compensation Cost Can Be Implemented ($F^{opt} \leq F^\circ$)

The following proposition states that, under the assumption that the optimal cooperative tariff level can be implemented, the optimal cooperative tariff level exceeds the optimal cooperative tariff level in the static game.

Proposition 13 *If the optimal compensation cost is applicable, the cooperative tariff level that maximizes the expected per-period payoff is strictly smaller than the optimal cooperative tariff level in the absence of uncertainty.*

Proof. Plugging the optimal compensation cost $\varepsilon_{EC}^{opt} = \frac{W_D - W_N}{W_D - W_C}$ into equation (11) renders the following per-period payoff:

$$W_E = W_N + [W_D - W_N]\Phi(\varepsilon_{EC}^{opt}) - [W_D - W_C] \int_0^{\varepsilon_{EC}^{opt}} \varepsilon d\Phi$$

The first derivative of W_E with regard to t_C is given by

$$\begin{aligned}
\frac{\partial W_E}{\partial t_C} &= \Phi(\varepsilon_{EC}^{opt}) \frac{\partial W_D}{\partial t_C} + [W_D - W_N] \varphi(\varepsilon_{EC}^{opt}) \frac{\partial \varepsilon_{EC}^{opt}}{\partial t_C} \\
&\quad - \int_0^{\varepsilon_{EC}^{opt}} \varepsilon d\Phi \frac{\partial [W_D - W_C]}{\partial t_C} - [W_D - W_C] \varphi(\varepsilon_{EC}^{opt}) \varepsilon_{EC}^{opt} \frac{\partial \varepsilon_{EC}^{opt}}{\partial t_C} \\
&= \Phi(\varepsilon_{EC}^{opt}) \frac{\partial W_D}{\partial t_C} - \int_0^{\varepsilon_{EC}^{opt}} \varepsilon d\Phi \frac{\partial [W_D - W_C]}{\partial t_C} \\
&\quad + [W_D - W_N] \varphi(\varepsilon_{EC}^{opt}) \left\{ 1 - \frac{W_D - W_C}{W_D - W_N} \varepsilon_{EC}^{opt} \right\} \frac{\partial \varepsilon_{EC}^{opt}}{\partial t_C} \\
&\stackrel{\varepsilon_{EC}^{opt} = \frac{W_D - W_N}{W_D - W_C}}{=} \Phi(\varepsilon_{EC}^{opt}) \frac{\partial W_D}{\partial t_C} - \int_0^{\varepsilon_{EC}^{opt}} \varepsilon d\Phi \frac{\partial [W_D - W_C]}{\partial t_C} \\
&= \Phi(\varepsilon_{EC}^{opt}) \frac{\partial W_D}{\partial t_C} - \Phi(\varepsilon_{EC}^{opt}) E(\varepsilon | \varepsilon \leq \varepsilon_{EC}^{opt}) \frac{\partial [W_D - W_C]}{\partial t_C} \\
&= \Phi(\varepsilon_{EC}^{opt}) \left\{ [1 - E(\varepsilon | \varepsilon \leq \varepsilon_{EC}^{opt})] \frac{\partial W_D}{\partial t_C} + E(\varepsilon | \varepsilon \leq \varepsilon_{EC}^{opt}) \frac{\partial W_C}{\partial t_C} \right\} \quad (16)
\end{aligned}$$

Since $\frac{\partial W_D}{\partial t_C} < 0$ for all t_C , it can easily be established that for any $t_C \geq t_C^{opt}$ (i.e. when $\frac{\partial W_C}{\partial t_C} \leq 0$) it is the case that $\frac{\partial W_E}{\partial t_C} < 0$. In particular, for the optimal static cooperative tariff level t_C^{opt} (i.e. when $\frac{\partial W_C}{\partial t_C} = 0$) it is the case that $\frac{\partial W_E}{\partial t_C} < 0$. Hence W_E is increasing below the static optimal cooperative tariff. ■

The first term of (16), $\Phi(\varepsilon_{EC}^{opt})$, is strictly positive for any cooperative tariff. Since the second term of (16) is strictly negative for $t_C = t_N$ it follows that the expected per-period payoff increases as the cooperative tariff level falls below the Nash tariff level.

The result of proposition 4 is somewhat surprising. How can setting the cooperative tariff level below the static optimal cooperative tariff increase the expected per-period payoff? It has to be kept in mind that a decline in t_C below t_C^{opt} leads to a fall in the expected outcome under mutual cooperation and to a fall in the expected outcome of being deviated against but also to an increase in the expected outcome under deviation. As long as the second effect outweighs the first two effects a further fall in the cooperative tariff is worthwhile. Theoretically one implication of proposition 4 could be that the expected per-period payoff continues to increase as the cooperative tariff decreases and eventually

goes to $-\infty$. This is however not the case, as the following proposition demonstrates.

Proposition 14 *There exists a finite solution to equation (16).*

Proof. Assume that there exists no finite solution to (16). Some rearrangement of equation (16) gives us the following optimality condition.

$$\begin{aligned}
\Phi(\varepsilon_{EC}^{opt}) \{ [1 - E(\varepsilon|\varepsilon \leq \varepsilon_{EC}^{opt})] \frac{\partial W_D}{\partial t_C} + E(\varepsilon|\varepsilon \leq \varepsilon_{EC}^{opt}) \frac{\partial W_C}{\partial t_C} \} &= 0 \\
\Leftrightarrow \frac{1 - E(\varepsilon|\varepsilon \leq \varepsilon_{EC}^{opt})}{E(\varepsilon|\varepsilon \leq \varepsilon_{EC}^{opt})} &= - \frac{\frac{\partial W_C}{\partial t_C}}{\frac{\partial W_D}{\partial t_C}} \\
\Leftrightarrow \frac{1}{E(\varepsilon|\varepsilon \leq \varepsilon_{EC}^{opt})} &= 1 - \frac{\frac{\partial W_C}{\partial t_C}}{\frac{\partial W_D}{\partial t_C}} \\
\Leftrightarrow E(\varepsilon|\varepsilon \leq \varepsilon_{EC}^{opt}) &= \frac{\frac{\partial W_D}{\partial t_C}}{\frac{\partial W_D}{\partial t_C} - \frac{\partial W_C}{\partial t_C}} = \frac{1}{1 - \frac{\frac{\partial W_C}{\partial t_C}}{\frac{\partial W_D}{\partial t_C}}} \quad (17)
\end{aligned}$$

From proposition 4 it follows that, because $\frac{\partial W_E}{\partial t_C} < 0$ for $t_C = t_C^{opt}$, W_E is maximized for some $t_C < t_C^{opt}$ (or it increases monotonically as $t_C \rightarrow -\infty$). For $t_C = t_N$ it is the case that $\frac{\frac{\partial W_C}{\partial t_C}}{\frac{\partial W_D}{\partial t_C}} = 1$. As t_C decreases $\frac{\frac{\partial W_C}{\partial t_C}}{\frac{\partial W_D}{\partial t_C}}$ decreases monotonously and goes to $-\infty$. Hence the right-hand side of (17) goes to infinity as t_C converges to t_N (from below) and it decreases as t_C decreases. For $t_C = t_C^{opt}$, i.e. when $\frac{\partial W_C}{\partial t_C} = 0$, its value is unity, and by letting t_C approach $-\infty$ it converges to zero. Since ε_{EC}^{opt} goes to infinity as t_C converges to t_N and decreases monotonously as t_C falls, the left-hand side equals unity for $t_C = t_N$, decreases as t_C falls and converges to zero. Thus for $t_C = t_C^{opt}$ the left-hand side is strictly smaller than the right-hand side. For there to exist no finite solution to (17) it has to be that the left-hand side is strictly smaller than the right-hand side for any $t_C < t_C^{opt}$. However letting t_C approach $-\infty$ the two sides will both converge to zero. Thus only by letting t_C go to $-\infty$ will equation (17) hold. This implies $\varepsilon_{EC}^{opt} = 0$, i.e. both countries will always deviate, in which case their expected payoff will be the Nash payoff, the same expected payoff that is achieved for $t_C = t_N$. From equation (16) it follows that the expected per-period payoff increases unambiguously as t_C decreases below t_N . Thus per-period payoff cannot be maximized by letting t_C approach $-\infty$, and there has to be a $t_C \in (-\infty, t_C^{opt})$ that maximizes the expected per-period payoff. ■

5.4.2 When the Optimal Compensation Cost Cannot Be Implemented ($F^{opt} > F^\circ$)

If the optimal cost for using the escape clause cannot be implemented, the highest possible cost F° will be implemented, and the escape clause will be used to the least possible degree. Using equations (9) and (11), an expression similar to the case in section 3 when there is no escape clause (cf. equation (7)) can be obtained for the case in which the maximum compensation cost F° is applied.

$$(9): \quad \varepsilon^\circ = \frac{\beta}{1-\beta} \frac{W_E - W_N}{W_D - W_C}$$

$$(11): \quad W_E - W_N = [W_D - W_N]\Phi(\varepsilon^\circ) - [W_D - W_C] \int_0^{\varepsilon^\circ} \varepsilon d\Phi$$

$$\Rightarrow \varepsilon^\circ = \frac{\beta}{1-\beta} \left\{ \frac{W_D - W_N}{W_D - W_C} \Phi(\varepsilon^\circ) - \int_0^{\varepsilon^\circ} \varepsilon d\Phi \right\} \equiv f^\circ(\varepsilon^\circ) \quad (18)$$

Proposition 15 *There exists at least one stable solution ($\varepsilon^\circ = 0$) to equation (18).*

Proof. (i) It is easily established that $f^\circ(0) = 0$. Hence $\varepsilon^\circ = 0$ is always a solution to (18).

(ii) By taking the first derivative of f° , it is possible to show that it is a stable solution.

$$\begin{aligned} \frac{\partial f^\circ(\varepsilon^\circ)}{\partial \varepsilon^\circ} &= \frac{\beta}{1-\beta} \left\{ \frac{W_D - W_N}{W_D - W_C} \varphi(\varepsilon^\circ) - \varepsilon^\circ \varphi(\varepsilon^\circ) \right\} \\ &= \frac{\beta}{1-\beta} \varphi(\varepsilon^\circ) \left\{ \frac{W_D - W_N}{W_D - W_C} - \varepsilon^\circ \right\} \end{aligned} \quad (19)$$

$\Rightarrow f^{\circ\prime}(0) = 0$ and $f^{\circ\prime}(\varepsilon^\circ) \geq 0$ and $\Phi(\varepsilon^\circ) < \varepsilon^\circ$ close to $\varepsilon^\circ = 0$ for any $t_C < t_N$

$\Rightarrow \varepsilon' = 0$ is a stable solution ■

Letting ε° go to infinity, f° converges to $\frac{\beta}{1-\beta} \frac{W_C - W_N}{W_D - W_C}$. There may or may not exist further solutions, depending on the discount factor and on the cooperative tariff.

By increasing the discount factor, $f^\circ(\varepsilon^\circ)$ increases for any given ε° . Since letting β approach unity will make the term $\frac{\beta}{1-\beta}$ very large, it is

the case that for a sufficiently high discount factor there will be values of ε° , for which $f^\circ(\varepsilon^\circ) \geq \varepsilon^\circ$. Similarly, by decreasing the discount factor, $f^\circ(\varepsilon^\circ)$ decreases for any given ε° , and eventually $f^\circ(\varepsilon^\circ) < \varepsilon^\circ$ for all $\varepsilon^\circ > 0$. Thus there exists a threshold value of β , below which there exist no further solutions for equation (18).

The reasoning is similar for the chosen cooperative tariff. The term $\frac{W_D - W_N}{W_D - W_C}$ goes to infinity as $t_C \rightarrow t_N$. By letting t_C decrease, this term falls and eventually goes to zero as $t_C \rightarrow -\infty$. Thus, by decreasing t_C , $f^\circ(\varepsilon^\circ)$ decreases for any given ε° . If t_C is sufficiently low, i.e. $\frac{W_D - W_N}{W_D - W_C}$ is sufficiently low, $f^\circ(\varepsilon^\circ) < \varepsilon^\circ$ for all $\varepsilon^\circ > 0$. In this case there will exist no further solution to equation (18). However for t_C sufficiently close to t_N there will at least be an interval of values for ε° , for which $f^\circ(\varepsilon^\circ) \geq \varepsilon^\circ$.

Proposition 16 *Among the solutions to (7), the highest will be chosen since it is both stable and renders the highest continuation value.*

Proof. (i) Since $f^\circ(\varepsilon^\circ)$ is bounded from above it follows immediately that for the highest value of ε° that solves (18) it must be true that $0 < f^{\circ\prime}(\varepsilon^\circ) \leq 1$. Thus this solution is stable.

(ii) By showing that for any solution of (18) the expected per-period payoff is increasing in ε° , it can be concluded that the highest value of ε° that solves (18) also renders the highest expected per-period payoff. Let $\varepsilon^{\circ\circ}$ be a solution to (18), i.e.

$$\begin{aligned} \varepsilon^{\circ\circ} = \Phi(\varepsilon^{\circ\circ}) &= \frac{\beta}{1 - \beta} \left\{ \frac{W_D - W_N}{W_D - W_C} \Phi(\varepsilon^{\circ\circ}) - \int_0^{\varepsilon^{\circ\circ}} \varepsilon d\Phi \right\} \\ &= \frac{\beta}{1 - \beta} \Phi(\varepsilon^{\circ\circ}) \left\{ \frac{W_D - W_N}{W_D - W_C} - E(\varepsilon | \varepsilon \leq \varepsilon^{\circ\circ}) \right\} \\ &\Leftrightarrow [W_D - W_N] - [W_D - W_C] E(\varepsilon | \varepsilon \leq \varepsilon^{\circ\circ}) = [W_D - W_C] \frac{(1 - \beta)\varepsilon^{\circ\circ}}{\beta\Phi(\varepsilon^{\circ\circ})} \end{aligned}$$

The derivative of the expected per-period payoff with respect to ε° for $\varepsilon^\circ = \varepsilon^{\circ\circ}$ is given by

$$\begin{aligned} \frac{\partial W_E}{\partial \varepsilon^\circ} \Big|_{\varepsilon^\circ = \varepsilon^{\circ\circ}} &= [W_D - W_N] \varphi(\varepsilon^{\circ\circ}) - [W_D - W_C] \varphi(\varepsilon^{\circ\circ}) \varepsilon^{\circ\circ} \\ &= \varphi(\varepsilon^{\circ\circ}) \{ [W_D - W_N] - \varepsilon^{\circ\circ} [W_D - W_C] \} \geq 0 \\ &\Leftrightarrow \varepsilon^{\circ\circ} \leq \frac{W_D - W_N}{W_D - W_C} = \varepsilon_{EC}^{opt} \end{aligned}$$

Since it is assumed that the optimal compensation cost cannot be implemented, i.e. $F^{opt} > F^\circ$, which implies $\varepsilon_{EC}^{opt} > \varepsilon^\circ$, it follows that for

any solution to (18) it is the case that the expected per-period payoff is increasing. It follows immediately that the highest value for ε° that solves (18) must be the one rendering the highest expected per-period payoff among the solutions. ■

By introspection of equation (19) it is easy to conclude that f° increases monotonously for $0 < \varepsilon^\circ < \frac{W_D - W_N}{W_D - W_C} = \varepsilon_{EC}^{opt}$ and decreases monotonously for $\varepsilon^\circ > \varepsilon_{EC}^{opt}$. Hence f° has a unique maximum at $\varepsilon^\circ = \varepsilon_{EC}^{opt}$. As ε° goes to infinity, f° converges to $\frac{\beta}{1-\beta} \frac{W_C - W_N}{W_D - W_C}$.

The expression for f° is similar to (7), and f° shares many features of f . For $\varepsilon = 0$ the values of f and f° coincide. As ε goes to infinity, f and f° converge to the same value. For any $0 < \varepsilon < \infty$ it is however the case that f° is strictly larger than f , as is demonstrated below.

$$\begin{aligned} f(\varepsilon') &= \frac{\beta\Phi(\varepsilon)}{1 - \beta\Phi^2(\varepsilon)} \left\{ \frac{W_D - W_N}{W_D - W_C} \Phi(\varepsilon) - \int_0^\varepsilon \varepsilon d\Phi \right\} \\ &< \frac{\beta}{1 - \beta} \left\{ \frac{W_D - W_N}{W_D - W_C} \Phi(\varepsilon) - \int_0^\varepsilon \varepsilon d\Phi \right\} = f^\circ(\varepsilon) \end{aligned}$$

Hence for any solution ε' of (7), except for the trivial one ($\varepsilon' = 0$), it is true that $f^\circ(\varepsilon') > \varepsilon'$. Thus given a certain discount factor and a certain cooperative tariff level and given that there exists more than one solution to (7), the solution of (18) with the highest value is strictly larger than the solution of (7) with the highest value. Moreover, since $f^\circ(\varepsilon') > f(\varepsilon')$ for $\varepsilon' > 0$, it is also the case that even if there exists no positive solution to (7), i.e. $f(\varepsilon') < \varepsilon'$ for $\varepsilon > 0$, there exist more than one solution of (18) for a sufficiently high discount factor and/or a sufficiently high cooperative tariff. Thus, given a certain discount factor, the range of cooperative tariffs, within which there exist strictly positive threshold value solutions, is wider when an escape clause is introduced. If however the discount factor is too low and/or the cooperative tariff level is too low, it will be the case that $f^\circ(\varepsilon') < \varepsilon'$ for $\varepsilon' > 0$ and hence there will exist no solution of (18) except the trivial one.

Conclusion 17 *When an escape clause with the maximal possible compensation cost is implemented (and a strictly positive solution for the threshold value exists) the threshold value above which the escape clause is exercised is strictly larger than the threshold value above which in the absence of an escape clause deviation will occur. Moreover, the range of the discount factor and the range of the cooperative tariff, for which*

strictly positive solutions for the threshold value exist, is strictly larger with than without an escape clause.

To conclude, given a certain discount factor, it is the case that there is a lower limit to the cooperative tariff level in order to obtain a strictly positive solution to (18). Thus the optimal cooperative tariff level has to be larger than this threshold cooperative tariff level. Next the properties of the optimal cooperative tariff level when the optimal compensation cost cannot be applied will be examined more closely.

Proposition 18 *The optimal cooperative tariff level when the maximal compensation cost is applied is higher than the optimal cooperative tariff level when the optimal compensation cost is applied.*

Proof. In case the optimal compensation cost cannot be applied, the highest possible compensation cost f° will be applied. The first-order condition for the optimal cooperative tariff level is given by

$$\begin{aligned} \frac{\partial W_E}{\partial t_C} &= \Phi(\varepsilon^\circ) \frac{\partial W_D}{\partial t_C} + [W_D - W_N] \varphi(\varepsilon^\circ) \frac{\partial \varepsilon^\circ}{\partial t_C} \\ &\quad - \int_0^{\varepsilon^\circ} \varepsilon d\Phi \frac{\partial [W_D - W_C]}{\partial t_C} - [W_D - W_C] \varphi(\varepsilon^\circ) \varepsilon^\circ \frac{\partial \varepsilon^\circ}{\partial t_C} \\ &= \Phi(\varepsilon^\circ) \frac{\partial W_D}{\partial t_C} - \int_0^{\varepsilon^\circ} \varepsilon d\Phi \frac{\partial [W_D - W_C]}{\partial t_C} \\ &\quad + [W_D - W_C] \varphi(\varepsilon^\circ) \left\{ \frac{W_D - W_N}{W_D - W_C} - \varepsilon^\circ \right\} \frac{\partial \varepsilon^\circ}{\partial t_C} \end{aligned}$$

Using equation (8) the derivative of the threshold value with respect to the cooperative tariff can be obtained.

$$\begin{aligned} \frac{\partial \varepsilon^\circ}{\partial t_C} &= \frac{\beta}{1 - \beta} \frac{1}{W_D - W_C} \left\{ \frac{\partial W_E}{\partial t_C} - \frac{W_E - W_N}{W_D - W_C} \frac{\partial [W_D - W_C]}{\partial t_C} \right\} \\ &= \frac{1}{W_D - W_C} \left\{ \frac{\beta}{1 - \beta} \frac{\partial W_E}{\partial t_C} - \varepsilon^\circ \frac{\partial [W_D - W_C]}{\partial t_C} \right\} \end{aligned}$$

Plugging this expression in the one above gives us

$$\begin{aligned}
& \frac{\partial W_E}{\partial t_C} \\
&= \Phi(\varepsilon^\circ) \frac{\partial W_D}{\partial t_C} - \int_0^{\varepsilon^\circ} \varepsilon d\Phi \frac{\partial [W_D - W_C]}{\partial t_C} \\
&+ \varphi(\varepsilon^\circ) \left[\frac{W_D - W_N}{W_D - W_C} - \varepsilon^\circ \right] \left[\frac{\beta}{1 - \beta} \frac{\partial W_E}{\partial t_C} - \varepsilon^\circ \frac{\partial [W_D - W_C]}{\partial t_C} \right] \\
&\Rightarrow \left\{ 1 - \frac{\beta}{1 - \beta} \varphi(\varepsilon^\circ) \left[\frac{W_D - W_N}{W_D - W_C} - \varepsilon^\circ \right] \right\} \frac{\partial W_E}{\partial t_C} \\
&= \Phi(\varepsilon^\circ) \frac{\partial W_D}{\partial t_C} - \int_0^{\varepsilon^\circ} \varepsilon d\Phi \frac{\partial [W_D - W_C]}{\partial t_C} \\
&\quad - \varphi(\varepsilon^\circ) \varepsilon^\circ \left[\frac{W_D - W_N}{W_D - W_C} - \varepsilon^\circ \right] \frac{\partial [W_D - W_C]}{\partial t_C} \\
&= \Phi(\varepsilon^\circ) \frac{\partial W_D}{\partial t_C} - \left\{ \int_0^{\varepsilon^\circ} \varepsilon d\Phi + \varphi(\varepsilon^\circ) \varepsilon^\circ \left[\frac{W_D - W_N}{W_D - W_C} - \varepsilon^\circ \right] \right\} \frac{\partial [W_D - W_C]}{\partial t_C} = 0 \\
&\Leftrightarrow \frac{\Phi(\varepsilon^\circ)}{\int_0^{\varepsilon^\circ} \varepsilon d\Phi + \varphi(\varepsilon^\circ) \varepsilon^\circ \left[\frac{W_D - W_N}{W_D - W_C} - \varepsilon^\circ \right]} = 1 - \frac{\frac{\partial W_C}{\partial t_C}}{\frac{\partial W_D}{\partial t_C}} \\
&\Leftrightarrow \frac{\int_0^{\varepsilon^\circ} \varepsilon d\Phi + \varphi(\varepsilon^\circ) \varepsilon^\circ \left[\frac{W_D - W_N}{W_D - W_C} - \varepsilon^\circ \right]}{\Phi(\varepsilon^\circ)} = \frac{1}{1 - \frac{\frac{\partial W_C}{\partial t_C}}{\frac{\partial W_D}{\partial t_C}}} \\
&\Leftrightarrow E(\varepsilon | \varepsilon \leq \varepsilon^\circ) + \frac{\varepsilon^\circ \varphi(\varepsilon^\circ)}{\Phi(\varepsilon^\circ)} \left\{ \frac{W_D - W_N}{W_D - W_C} - \varepsilon^\circ \right\} = \frac{1}{1 - \frac{\frac{\partial W_C}{\partial t_C}}{\frac{\partial W_D}{\partial t_C}}} \quad (20)
\end{aligned}$$

Next, it is shown that the left-hand side of (20) is larger than the left-hand side of (17).

$$\begin{aligned}
& \int_0^{\varepsilon^\circ} \varepsilon d\Phi + \varepsilon^\circ \varphi(\varepsilon^\circ) \left\{ \frac{W_D - W_N}{W_D - W_C} - \varepsilon^\circ \right\} \\
&= \int_0^{\varepsilon_{EC}} \varepsilon d\Phi - \int_{\varepsilon^\circ}^{\varepsilon_{EC}} \varepsilon d\Phi + \varepsilon^\circ \varphi(\varepsilon^\circ) [\varepsilon_{EC}^{opt} - \varepsilon^\circ] \\
&\geq \int_0^{\varepsilon_{EC}} \varepsilon d\Phi - \varepsilon^\circ \varphi(\varepsilon^\circ) [\varepsilon_{EC}^{opt} - \varepsilon^\circ] + \varepsilon^\circ \varphi(\varepsilon^\circ) [\varepsilon_{EC}^{opt} - \varepsilon^\circ] \\
&= \int_0^{\varepsilon_{EC}} \varepsilon d\Phi \\
&\Rightarrow \frac{\int_0^{\varepsilon^\circ} \varepsilon d\Phi + \varphi(\varepsilon^\circ) \varepsilon^\circ \left[\frac{W_D - W_N}{W_D - W_C} - \varepsilon^\circ \right]}{\Phi(\varepsilon^\circ)} \geq \frac{\int_0^{\varepsilon_{EC}} \varepsilon d\Phi}{\Phi(\varepsilon_{EC}^{opt})}
\end{aligned}$$

When $\varepsilon^\circ < \varepsilon_{EC}^{opt}$, the left-hand side of (20) is strictly larger than the left-hand side of (17). Thus for any $t_C < t_N$ the left-hand side of (20) attains a larger value than the left-hand side in (17), and hence the intersection with the right-hand side has to be to the left, i.e. at a higher cooperative tariff level, in (20) compared to (17). ■

(Alternative proof: Since it is assumed that the optimal compensation cost is not applicable, it follows immediately that $f^\circ(\varepsilon_{EC}^{opt}) \leq \varepsilon_{EC}^{opt}$. Since f° is maximized for $\varepsilon^\circ = \varepsilon_{EC}^{opt}$, the following must be true for the maximal solution to (18): $\varepsilon^{\circ max} = f^\circ(\varepsilon^{\circ max}) \leq f^\circ(\varepsilon_{EC}^{opt}) \leq \varepsilon_{EC}^{opt}$)

Proposition 19 *The optimal cooperative tariff level when the maximal compensation cost is applied decreases in the discount factor.*

Proof. Equation (12) implies that an increase in b leads to an increase in ε° , which in turn leads to an increase in W_E as long as $\varepsilon^\circ < \varepsilon_{EC}^{opt}$ in (14). The increase in W_E reinforces the increase in ε° in (9). Next the impact of an increase in ε° on the left-hand side of (20) is evaluated.

$$E(\varepsilon | \varepsilon \leq \varepsilon^\circ) + \frac{\varphi(\varepsilon^\circ) \varepsilon^\circ}{\Phi(\varepsilon^\circ)} \left[\frac{W_D - W_N}{W_D - W_C} - \varepsilon^\circ \right] = \frac{\int_0^{\varepsilon^\circ} \varepsilon d\Phi + \varepsilon^\circ \varphi(\varepsilon^\circ) \left[\frac{W_D - W_N}{W_D - W_C} - \varepsilon^\circ \right]}{\Phi(\varepsilon^\circ)}$$

The denominator obviously increases in ε° . To assess the effect of an increase in ε° on the nominator it is necessary to take the first derivative.

$$\begin{aligned} & \varepsilon^\circ \varphi(\varepsilon^\circ) + [\varphi(\varepsilon^\circ) + \varepsilon^\circ \varphi'(\varepsilon^\circ)] \left[\frac{W_D - W_N}{W_D - W_C} - \varepsilon^\circ \right] - \varphi(\varepsilon^\circ) \varepsilon^\circ \\ & = [\varphi(\varepsilon^\circ) + \varepsilon^\circ \varphi'(\varepsilon^\circ)] \left[\frac{W_D - W_N}{W_D - W_C} - \varepsilon^\circ \right] \end{aligned}$$

Since $\varphi(\varepsilon^\circ) = \int_0^{\varepsilon^\circ} \varphi'(\varepsilon) d\varepsilon \leq \varepsilon^\circ \varphi'(\varepsilon^\circ)$, it follows immediately that the first derivative is negative, as long as $\varepsilon^\circ < \varepsilon_{EC}^{opt}$. Hence an increase in ε° leads unambiguously to a decrease in the left-hand side of (20). This implies that the cooperative tariff level that solves (20) has to be lower. Thus it can be concluded that an increase in b leads to a lower optimal cooperative tariff level. ■

5.5 The Gains From Introducing an Escape Clause

By introducing an escape clause it can be avoided that cooperation breaks down. It is however also the case that the inclusion of an escape clause increases the expected payoff even if the optimal compensation cost cannot be implemented.

Proposition 20 *Given a cooperative tariff strictly smaller than the Nash tariff and that the optimal compensation cost cannot be applied, the threshold level, above which the escape clause is used, is strictly larger than the threshold level, above which cooperation breaks down in the absence of an escape clause ($\varepsilon^\circ > \varepsilon'$), as long as it is strictly positive ($\varepsilon^\circ > 0$). Moreover, even though the optimal compensation cost cannot be applied, the continuation value of the game exceeds the continuation value in the absence of an escape clause ($V'_C > (1 - \beta)V_C$).*

Proof. (i) For very low discount factors it is the case that $\varepsilon^\circ = \varepsilon' = 0$. For a range of higher discount factors it is the case that $\varepsilon' = 0$ but $\varepsilon^\circ > 0$, and the result is trivial. For sufficiently high discount factors ε' is strictly positive and is defined by (7), while ε° is defined by (10).

$$\begin{aligned} \varepsilon^\circ = f^\circ(\varepsilon^\circ) &= \frac{\beta}{1 - \beta} \left\{ \frac{W_D - W_N}{W_D - W_C} \Phi(\varepsilon^\circ) - \int_0^{\varepsilon^\circ} \varepsilon d\Phi \right\} \\ &> \frac{\beta \Phi(\varepsilon^\circ)}{1 - \beta \Phi^2(\varepsilon^\circ)} \left\{ \frac{W_D - W_N}{W_D - W_C} \Phi(\varepsilon^\circ) - \int_0^{\varepsilon^\circ} \varepsilon d\Phi \right\} = f(\varepsilon^\circ) \\ &\Rightarrow \varepsilon^\circ < \varepsilon' \end{aligned}$$

(ii)

$$\begin{aligned}
& (1 - \beta)V_C - W_N \\
&= \frac{1 - \beta}{1 - \beta\Phi^2(\varepsilon')} \{ [W_D - W_N]\Phi(\varepsilon') - [W_D - W_C] \int_0^{\varepsilon'} \varepsilon d\Phi \} \\
&< [W_D - W_N]\Phi(\varepsilon') - [W_D - W_C] \int_0^{\varepsilon'} \varepsilon d\Phi \\
&< [W_D - W_N]\Phi(\varepsilon^\circ) - [W_D - W_C] \int_0^{\varepsilon^\circ} \varepsilon d\Phi \quad \text{since } \varepsilon' < \varepsilon^\circ \leq \varepsilon_{EC}^{opt} = \frac{W_D - W_N}{W_D - W_C} \\
&= W_E - W_N \\
&\Rightarrow V_C < \frac{W_E}{1 - \beta} = V_C'
\end{aligned}$$

Thus the continuation value of the game with an escape clause and the maximum compensation cost for using it is strictly larger than the continuation value of the game without an escape clause. ■

The following lemma is straightforward.

Lemma 21 *The discounted expected payoff when there is an escape clause and the optimal compensation cost can be applied is strictly larger than the continuation value in the absence of an escape clause mechanism.*

Proof. Applying the optimal compensation cost yields the highest possible expected per-period payoff. Thus the result follows immediately from the previous proposition. ■

Thus, except helping to sustain an agreement, there is a further benefit from introducing an escape clause. It increases the expected payoff under the agreement.

6 Conclusions

Introducing uncertain outcomes into a repeated Prisoner's Dilemma tariff setting game will lead to breakdown of cooperation for sufficiently strong shocks. By introducing an escape clause mechanism that allows for temporary deviation from cooperation it is possible to always sustain cooperative behavior even as shocks that increase the one-period

gain from deviating occur. Moreover, the expected per-period payoff increases.

For a sufficiently high discount factor, the threshold level for temporary deviation through the escape clause will be at the efficient level. In this case the optimal cooperative tariff is strictly smaller than the optimal cooperative tariff when there is no uncertainty.

The threshold level, above which the escape clause is exercised, decreases as the cooperative tariff falls. A lower cooperative tariff is thus associated with an increase in the use of the escape clause.

7 References

- Abreu, Dilip, Pearce, David, and Stacchetti, Ennio (1986), *Optimal Cartel Equilibria with Imperfect Monitoring*, Journal of Economic Theory, 39, 251-269
- Bagwell, Kyle and Staiger, Robert (1990), *A Theory of Managed Trade*, American Economic Review, 80(4), 779-795
- Bagwell, Kyle and Staiger, Robert (2001), *The Economics of the World Trading System*, mimeo
- Bütler, Monika and Hauser, Heinz (2000), *The WTO Dispute Settlement System: A First Assessment from an Economic Perspective*, Journal of Law, Economics and Organization, 16(2), 503-533
- Dixit, Avinash (1987), *Strategic Aspects of Trade Policy*, in Bewley, T.F. (ed.), *Advances in Economic Theory: Proceedings of the Fifth World Congress of the Economic Society*, 329-362, Cambridge University Press, Cambridge, UK
- Ethier, Wilfred J. (1998), *The International Commercial System*, Essays in International Finance No. 210, International Finance Section, Department of Economics, Princeton University, Princeton, New Jersey
- Ethier, Wilfried J. (2001), *Theoretical Problems in Negotiating Trade Liberalization*, European Journal of Political Economy, 17, 209-232
- Green, Edward J. and Porter, Robert H. (1984), *Noncooperative Collusion under Imperfect Price Information*, Econometrica, 52(1), 87-100
- Hoekman, Bernard M. and Kostecki (1995), Michael M., *The Political Economy of the World Trading System*, Oxford University Press, Oxford, UK
- Hungerford, Thomas L. (1991), *GATT: A Cooperative Equilibrium in a Noncooperative Trading Regime*, Journal of International Economics, 31, 357-369
- Kovenock, Dan and Thursby, Marie (1992), *GATT: Dispute Settlement and Cooperation*, Economics and Politics, 4(2), 151-170
- Ludema, Rodney D. (2001), *Optimal International Trade Agreements and Dispute Settlement Procedures*, European Journal of Political Economy, 17(2), 355-376

- Maggi, Giovanni (1999), *The Role of Multilateral Institutions in International Trade Cooperation*, American Economic Review, 89(1), 190-214
- Mayer, Wolfgang (1981), *Theoretical Considerations on Negotiated Tariff Adjustments*, Oxford Economic Papers, 33(1), 135-153
- Milgrom, Paul R., North, Douglass C., and Weingast, Barry R. (1990), *The Role of Institutions in the Revival of Trade: The Law Merchant, Private Judges, and the Champagne Fairs*, Economics and Politics, 2(1), 1-23
- Riezman, Raymond (1982), *Tariff Retaliation from a Strategic Viewpoint*, Southern Economic Journal, 48(3), 583-593
- Riezman, Raymond (1991), *Dynamic Tariffs with Asymmetric Information*, Journal of International Economics, 30(3-4), 267-284
- Rosendorff, B. Peter and Milner, Helen (2001), *The Optimal Design of International Trade Institutions: Uncertainty and Escape*, International Organization, 55(4), 829-857
- Schott, Jeffrey J. (1994), *The Uruguay Round: An Assessment*, Institute for International Economics, Washington, DC, USA
- World Trade Organization, *The General Agreement on Tariffs and Trade*, Geneva, Switzerland
- World Trade Organization, *Agreement on Safeguards*, Geneva, Switzerland

8 Appendix

8.1 Derivation of V_C

The continuation value depends on the probability of the trading partner opting for cooperation (p) and all possible realizations of the domestic shock ε .

$$\begin{aligned}
V_C &= p \left\{ \int_0^{\varepsilon'} [W_C^\varepsilon + \beta V_C] d\Phi + \int_{\varepsilon'}^\infty [W_D^\varepsilon + \beta V_D] d\Phi \right\} \\
&+ (1-p) \left\{ \int_0^{\varepsilon'} [W_S^\varepsilon + \beta V_D] d\Phi + \int_{\varepsilon'}^\infty [W_N^\varepsilon + \beta V_D] d\Phi \right\} \\
&= p \left\{ \int_0^{\varepsilon'} [W_D^\varepsilon - \varepsilon [W_D - W_C] + \beta V_C] d\Phi + \int_{\varepsilon'}^\infty [W_D^\varepsilon + \beta V_D] d\Phi \right\} \\
&+ (1-p) \left\{ \int_0^{\varepsilon'} [W_N^\varepsilon - \varepsilon [W_D - W_C] + \beta V_D] d\Phi + \int_{\varepsilon'}^\infty [W_N^\varepsilon + \beta V_D] d\Phi \right\} \quad \text{by (A3)} \\
&= p \left\{ \int_0^\infty W_D^\varepsilon d\Phi - \int_0^{\varepsilon'} \varepsilon [W_D - W_C] d\Phi + \int_0^{\varepsilon'} \beta V_C d\Phi + \int_{\varepsilon'}^\infty \beta V_D d\Phi \right\} \\
&+ (1-p) \left\{ \int_0^\infty [W_N^\varepsilon + \beta V_D] d\Phi - \int_0^{\varepsilon'} \varepsilon [W_D - W_C] d\Phi \right\} \\
&= p \left\{ W_D - [W_D - W_C] \int_0^{\varepsilon'} \varepsilon d\Phi + \beta p V_C + \beta (1-p) V_D \right\} \\
&+ (1-p) \left\{ W_N + \beta V_D - [W_D - W_C] \int_0^{\varepsilon'} \varepsilon d\Phi \right\} \\
&= p W_D + (1-p) W_N - [W_D - W_C] \int_0^{\varepsilon'} \varepsilon d\Phi + \beta p^2 V_C + \beta (1-p^2) V_D
\end{aligned}$$

$$\begin{aligned}
&\Leftrightarrow [1 - \beta p^2]V_C = pW_D + (1 - p)W_N - [W_D - W_C] \int_0^{\varepsilon'} \varepsilon d\Phi + \beta(1 - p^2)V_D \\
&= pW_D + (1 - \beta)(1 - p)V_D - [W_D - W_C] \int_0^{\varepsilon'} \varepsilon d\Phi + \beta(1 - p^2)V_D \quad \text{by (4)} \\
&= pW_D - [W_D - W_C] \int_0^{\varepsilon'} \varepsilon d\Phi + (1 - p)(1 + \beta p)V_D \\
&\Leftrightarrow V_C = \frac{1}{1 - \beta p^2} \{pW_D - [W_D - W_C] \int_0^{\varepsilon'} \varepsilon d\Phi + (1 - p)(1 + \beta p)V_D\} \\
&= V_D + \frac{1}{1 - \beta p^2} \{pW_D - [W_D - W_C] \int_0^{\varepsilon'} \varepsilon d\Phi - (1 - \beta)pV_D\} \\
&= V_D + \frac{1}{1 - \beta p^2} \{pW_D - [W_D - W_C] \int_0^{\varepsilon'} \varepsilon d\Phi - pW_N\} \quad \text{by (4)} \\
&= V_D + \frac{1}{1 - \beta p^2} \{p[W_D - W_N] - [W_D - W_C] \int_0^{\varepsilon'} \varepsilon d\Phi\} \\
&= V_D + \frac{1}{1 - \beta \Phi^2(\varepsilon')} \{[W_D - W_N]\Phi(\varepsilon') - [W_D - W_C] \int_0^{\varepsilon'} \varepsilon d\Phi\} \quad (5)
\end{aligned}$$

8.2 Proof of proposition 2

(i) Since $f(\varepsilon')$ is bounded from above it follows immediately that for the highest value of ε' that solves (7) it must be true that $0 < f'(\varepsilon') \leq 1$. Thus this solution is stable.

(ii) By showing that for any solution of (7) the continuation value is increasing in ε' , it can be concluded that the highest value of ε' that solves (7) also renders the highest continuation value. Let ε'' be a solu-

tion to (7), i.e.

$$\begin{aligned}\varepsilon'' = f(\varepsilon'') &= \frac{\beta\Phi^2(\varepsilon'')}{1 - \beta\Phi^2(\varepsilon'')} \left\{ \frac{W_D - W_N}{W_D - W_C} - E(\varepsilon|\varepsilon \leq \varepsilon'') \right\} \\ \Leftrightarrow [W_D - W_N] - [W_D - W_C]E(\varepsilon|\varepsilon \leq \varepsilon'') &= [W_D - W_C] \frac{[1 - \beta\Phi^2(\varepsilon'')]\varepsilon''}{\beta\Phi^2(\varepsilon'')}\end{aligned}$$

The derivative of the continuation value with respect to ε' for $\varepsilon' = \varepsilon''$ is given by

$$\begin{aligned}\frac{\partial V_C}{\partial \varepsilon'} \Big|_{\varepsilon' = \varepsilon''} &= \frac{1}{1 - \beta\Phi^2(\varepsilon'')} \{ [W_D - W_N]\varphi(\varepsilon'') - [W_D - W_C]\varphi(\varepsilon'')\varepsilon'' \} \\ &+ \frac{2\beta\Phi(\varepsilon'')\varphi(\varepsilon'')}{[1 - \beta\Phi^2(\varepsilon'')]^2} \{ [W_D - W_N]\Phi(\varepsilon'') - [W_D - W_C] \int_0^{\varepsilon''} \varepsilon d\Phi \} \\ &= \frac{\varphi(\varepsilon'')}{1 - \beta\Phi^2(\varepsilon'')} \{ [W_D - W_N] - [W_D - W_C]\varepsilon'' \} \\ &+ \frac{2\beta\Phi^2(\varepsilon'')\varphi(\varepsilon'')}{1 - \beta\Phi^2(\varepsilon'')} \{ [W_D - W_N] - [W_D - W_C]E(\varepsilon|\varepsilon \leq \varepsilon'') \} \\ &= \frac{\varphi(\varepsilon'')}{[1 - \beta\Phi^2(\varepsilon'')]^2} \{ \{ [W_D - W_N] - [W_D - W_C]\varepsilon'' \} [1 - \beta\Phi^2(\varepsilon'')] \} \\ &+ 2\beta\Phi^2(\varepsilon'') \{ [W_D - W_N] - [W_D - W_C]E(\varepsilon|\varepsilon \leq \varepsilon'') \} \\ &= \frac{\varphi(\varepsilon'')}{[1 - \beta\Phi^2(\varepsilon'')]^2} \{ \{ [W_D - W_N] - [W_D - W_C]\varepsilon'' \} [1 - \beta\Phi^2(\varepsilon'')] \} \\ &+ 2\beta\Phi^2(\varepsilon'')[W_D - W_C][1 - \beta\Phi^2(\varepsilon'')] \frac{\varepsilon''}{\beta\Phi^2(\varepsilon'')} \} \\ &= \frac{\varphi(\varepsilon'')}{1 - \beta\Phi^2(\varepsilon'')} \{ [W_D - W_N] + [W_D - W_C]\varepsilon'' \} \geq 0\end{aligned}$$

Hence for any solution to (7) it is the case that the continuation value is increasing. It follows immediately that the highest value for ε' that solves (7) must be the one rendering the highest continuation value among the solutions.

8.3 Derivation of W_E

The expected per-period payoff depends on the probability of the trading partner choosing cooperation, $p_C = \Phi(\varepsilon_{EC})$, and all possible realizations

of the domestic shock ε .

$$\begin{aligned}
W_E &= \Phi(\varepsilon_{EC}) \left\{ \int_0^{\varepsilon_{EC}} W_C^\varepsilon d\Phi + \int_{\varepsilon_{EC}}^\infty [W_D^\varepsilon - F] d\Phi \right\} \\
&+ [1 - \Phi(\varepsilon_{EC})] \left\{ \int_0^{\varepsilon_{EC}} [W_S^\varepsilon + F] d\Phi + \int_{\varepsilon_{EC}}^\infty W_N^\varepsilon d\Phi \right\} \\
&= \Phi(\varepsilon_{EC}) \left\{ \int_0^{\varepsilon_{EC}} [W_D^\varepsilon - \varepsilon[W_D - W_C]] d\Phi + \int_{\varepsilon_{EC}}^\infty [W_D^\varepsilon - F] d\Phi \right\} \\
&+ [1 - \Phi(\varepsilon_{EC})] \left\{ \int_0^{\varepsilon_{EC}} [W_N^\varepsilon - \varepsilon[W_N - W_S] + F] d\Phi + \int_{\varepsilon_{EC}}^\infty W_N^\varepsilon d\Phi \right\} \\
&= \Phi(\varepsilon_{EC}) \left\{ \int_0^\infty W_D^\varepsilon d\Phi - [W_D - W_C] \int_0^{\varepsilon_{EC}} \varepsilon d\Phi - \int_{\varepsilon_{EC}}^\infty F d\Phi \right\} \\
&+ [1 - \Phi(\varepsilon_{EC})] \left\{ \int_0^\infty W_N^\varepsilon d\Phi - [W_N - W_S] \int_0^{\varepsilon_{EC}} \varepsilon d\Phi + \int_0^{\varepsilon_{EC}} F d\Phi \right\} \\
&= \Phi(\varepsilon_{EC}) \left\{ W_D - [W_D - W_C] \int_0^{\varepsilon_{EC}} \varepsilon d\Phi - [1 - \Phi(\varepsilon_{EC})] F \right\} \\
&+ [1 - \Phi(\varepsilon_{EC})] \left\{ W_N - [W_D - W_C] \int_0^{\varepsilon_{EC}} \varepsilon d\Phi + \Phi(\varepsilon_{EC}) F \right\} \\
&= \Phi(\varepsilon_{EC}) W_D + [1 - \Phi(\varepsilon_{EC})] W_N - [W_D - W_C] \int_0^{\varepsilon_{EC}} \varepsilon d\Phi \\
&= W_N + \Phi(\varepsilon_{EC}) [W_D - W_N] - [W_D - W_C] \int_0^{\varepsilon_{EC}} \varepsilon d\Phi \tag{11}
\end{aligned}$$

8.4 Derivation of β°

$$\begin{aligned}
F^{opt} \leq F^\circ &\Leftrightarrow W_D - W_N \leq \frac{\beta}{1-\beta}[W_E - W_N] \\
&= \frac{\beta}{1-\beta}\{[W_D - W_N]\Phi(\varepsilon_{EC}^{opt}) - [W_D - W_C] \int_0^{\varepsilon_{EC}^{opt}} \varepsilon d\Phi\} \quad \text{by (13)} \\
&\Leftrightarrow (1-\beta)[W_D - W_N] \leq \beta\{[W_D - W_N]\Phi(\varepsilon_{EC}^{opt}) - [W_D - W_C] \int_0^{\varepsilon_{EC}^{opt}} \varepsilon d\Phi\} \\
&\Leftrightarrow W_D - W_N \leq \beta\{[1 + \Phi(\varepsilon_{EC}^{opt})][W_D - W_N] - [W_D - W_C] \int_0^{\varepsilon_{EC}^{opt}} \varepsilon d\Phi\} \\
&\Leftrightarrow \beta \geq \frac{W_D - W_N}{[1 + \Phi(\varepsilon_{EC}^{opt})][W_D - W_N] - [W_D - W_C] \int_0^{\varepsilon_{EC}^{opt}} \varepsilon d\Phi} \\
&= \frac{1}{1 + \Phi(\varepsilon_{EC}^{opt}) - \frac{W_D - W_C}{W_D - W_N} \int_0^{\varepsilon_{EC}^{opt}} \varepsilon d\Phi} \\
&= \frac{1}{1 + \Phi(\varepsilon_{EC}^{opt}) - \frac{\int_0^{\varepsilon_{EC}^{opt}} \varepsilon d\Phi}{\varepsilon_{EC}^{opt}}} \\
&= \frac{1}{1 + \Phi(\varepsilon_{EC}^{opt})[1 - \frac{E(\varepsilon|\varepsilon \leq \varepsilon_{EC}^{opt})}{\varepsilon_{EC}^{opt}}]} \equiv \beta^\circ \tag{14}
\end{aligned}$$

8.5 Derivation of equation (15)

$$\begin{aligned}
 \beta &= \frac{1}{\frac{\int_0^{\varepsilon_{EC}^{opt}} \varepsilon d\Phi}{\varepsilon_{EC}^{opt}} + \Phi(\varepsilon_{EC}^{opt})} \\
 &\Leftrightarrow 1 + \Phi(\varepsilon_{EC}^{opt}) - \frac{\int_0^{\varepsilon_{EC}^{opt}} \varepsilon d\Phi}{\varepsilon_{EC}^{opt}} = \frac{1}{\beta} \\
 &\Leftrightarrow 1 - \frac{1}{\beta} + \Phi(\varepsilon_{EC}^{opt}) - \frac{\int_0^{\varepsilon_{EC}^{opt}} \varepsilon d\Phi}{\varepsilon_{EC}^{opt}} = 0 \\
 &\Leftrightarrow \Phi(\varepsilon_{EC}^{opt}) - \frac{\int_0^{\varepsilon_{EC}^{opt}} \varepsilon d\Phi}{\varepsilon_{EC}^{opt}} = \frac{1}{\beta} - 1 \\
 &\Leftrightarrow \Phi(\varepsilon_{EC}^{opt}) \left[1 - \frac{E(\varepsilon | \varepsilon \leq \varepsilon_{EC}^{opt})}{\varepsilon_{EC}^{opt}} \right] = \frac{1 - \beta}{\beta} \tag{15}
 \end{aligned}$$