Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Journal of Environmental Economics and Management 51 (2006) 93–109
JOURNAL OF ENVIRONMENTAL ECONOMICS AND MANAGEMENT
www.elsevier.com/locate/jeem

Regional versus global cooperation for climate control

Geir B. Asheim[a], Camilla Bretteville Froyn[b,*], Jon Hovi[b,c], Fredric C. Menz[b,d]

[a]Department of Economics, University of Oslo, P.O. Box 1095 Blindern, 0317 Oslo, Norway
[b]Center for International Climate and Environmental Research – Oslo (CICERO), P.O. Box 1129 Blindern, 0318 Oslo, Norway
[c]Department of Political Science, University of Oslo, P.O. Box 1097 Blindern, 0317 Oslo, Norway
[d]Faculty of Economics and Finance, School of Business, Clarkson University, Potsdam, NY 13699, USA

Received 22 April 2003
Available online 22 July 2005

Abstract

This paper considers whether international environmental public goods provision, such as mitigation of climate change, is better dealt with through regional cooperation than through a global treaty. Previous research suggests that, at best, a global environmental treaty will achieve very little. At worst, it will fail to enter into force. Using a simple dynamic game-theoretic model, with weakly renegotiation-proof equilibrium as solution concept, we demonstrate that two agreements can sustain a larger number of cooperating parties than a single global treaty. The model provides upper and lower bounds on the number of parties under each type of regime. It is shown that a regime with two agreements can Pareto dominate a regime based on a single global treaty. We conclude that regional cooperation might be a good alternative–or supplement–to global environmental agreements.
© 2005 Elsevier Inc. All rights reserved.

Keywords: Climate change; International environmental agreements; Regional cooperation; The Kyoto Protocol; Non-cooperative game theory; Public goods; Weak renegotiation proofness

*Corresponding author. CICERO, P.O. Box 1129 Blindern, 0318 Oslo, Norway, Fax: +47 22 85 87 51.
E-mail addresses: g.b.asheim@econ.uio.no (G.B. Asheim), cbf@cicero.uio.no (C.B. Froyn), jon.hovi@stv.uio.no (J. Hovi), menzf@clarkson.edu (F.C. Menz).

0095-0696/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.jeem.2005.04.004

1. Introduction

To be effective, international environmental agreements (IEAs) need to achieve broad participation. For example, no international regime on climate change will be fully effective without the involvement of countries that have not yet committed to reduce emissions levels under the Kyoto framework, such as China, India, other developing countries, the United States, and Australia.

Previous research based on non-cooperative game theory suggests that, at best, a global IEA might achieve very little and at worst, could fail to enter into force (e.g., [5,7,14]). It is therefore interesting to consider whether two or more regional treaties might be more successful than a single global treaty. In this paper we show that an arrangement with two regional agreements can sustain a larger number of cooperating parties than a single global treaty. Therefore, in a second-best world characterized by free-rider incentives, a global treaty requiring agreement among a large number of countries is not necessarily the best option.

Regional cooperation to deal with climate change is already taking place. In March 2000, the European Climate Change Programme was launched by the European Union to coordinate EU efforts to reduce greenhouse gas emissions, including an emissions trading scheme within the EU that was implemented in 2005. At the same time there is speculation about possible U.S. interest in a regional climate arrangement that would include NAFTA members (Canada, Mexico, and the United States) and could be facilitated by the North American Commission on Environmental Cooperation. Given the possibility of at least two regional climate policy arrangements and the likelihood of more in the future as developing countries commit to reduce greenhouse gas emissions, the next phase of global cooperation for climate control could evolve into a number of regional alliances.

Our formal analysis is performed within the context of a simple dynamic game-theoretic model. We show that two agreements can sustain a larger number of cooperating parties than a global treaty even though the model assumes that the cost of reducing emissions is the same with both types of regime. The model provides upper and lower bounds on the number of parties under each regime. It is shown that a system with two agreements can Pareto-dominate a regime based on a global treaty. Thus, regional cooperation might be a good alternative to global IEAs for environmental problems like climate change.

The paper is organized as follows. Section 2 provides background information and reviews the existing literature on the conditions for an effective global IEA. Section 3 compares the conditions for a single global treaty to the conditions for two regional agreements in a simple dynamic model. Finally, Section 4 offers a summary and a discussion of the main results.

2. Background and literature review

There are IEAs designed to provide pure public goods on a global scale. For example, this is the case for the Kyoto protocol. Each country's emissions of greenhouse gases add to the world's atmospheric stock and a single country's efforts to control emissions will benefit all countries in a non-exclusive and non-rival manner. All parties thus have an incentive to free ride on other countries' efforts to reduce greenhouse gas emissions.

There are two types of incentive for free riding: the incentive for a country to not sign the agreement and thus benefit from the signatories' abatement efforts; and the incentive for a signatory to not comply with commitments made in an agreement. Because of these free-rider incentives, there will generally be suboptimal equilibrium coalition structures in global pollution control [22].

There is a considerable body of literature addressing under-provision of global international pollution control.[1] An important focus of this literature is the conditions leading to the formation of multilateral agreements or coalitions. Such agreements must be self-enforcing since there is no supranational authority to enforce compliance (e.g., [3,7]).

Models of international environmental cooperation differ with respect to the specification of the utility functions of governments and the stability concept they employ. However, they can roughly be divided into two groups — reduced-stage game models and dynamic game models [22].

Reduced-stage game models depict coalition formation as a two-stage game. In the first stage, countries decide on coalition formation. In the second stage, they choose abatement levels and how the gains from cooperation will be distributed (e.g., [18,14,23]).[2] A number of reduced-stage models have addressed the possibility of giving countries the freedom to negotiate more than one climate agreement [10–12,17]. A two-stage coalition game is used to show that when more than one coalition is possible, the equilibrium coalition structure that endogenously emerges from the negotiation process is characterized by several coalitions. It has also been shown that social welfare can be higher with multiple agreements than with a single global accord [12].

Dynamic game models typically assume an infinitely repeated game where governments agree on a contract in the first stage that has to be enforced in subsequent stages through credible threats (e.g., [3,5,7]).[3] The main problem analyzed by these models is that a country can — at least temporarily — achieve net gains by free riding or by joining a smaller coalition.

The literature on dynamic game models emphasizes that, to be effective, a treaty needs to specify a strategy that can enforce compliance. This entails that it must be in the (individual) best interest of each party to act in accordance with this strategy. Furthermore, such an equilibrium agreement must be renegotiation-proof. In particular, it must be in the (collective) best interest of other countries to insist that a non-compliant country be punished before cooperation can be resumed [5,7,22]. Main results from dynamic game models include that the allocation of abatement burdens crucially affects the success of international environmental agreements, that a grand coalition is unlikely to form, and that a sub-coalition may achieve more for its members than the grand coalition.

Barrett [6] uses a repeated game framework to demonstrate that a global treaty can be broadened to incorporate all countries (a consensus treaty), but at the cost of limiting the per-country level (the 'depth') of cooperation.[4] Below we show that regional agreements can enhance participation even without limiting the depth of cooperation. Our model largely follows Barrett [5], using Farrell and Maskin's [19] weak renegotiation-proof equilibrium as solution

[1]Notable examples are [3,5,7,13,14,23,24].

[2]For overviews of the coalition literature, see [10,20].

[3]Note that it is generally easier to sustain cooperation in infinitely repeated games than in finitely repeated games.

[4]Finus and Rundshagen [21] generalize Barrett's [3] model by introducing non-linear (quadratic) cost and benefit functions. In an appendix to this paper we consider how the impact of a regime based on regional agreements is influenced by introducing non-linear benefit functions.

concept.[5] The results support the conclusions reached by Carraro and others (e.g., [11,12,17]) using a reduced-stage game framework.

3. A simple dynamic model of regional climate agreements

Consider a world containing N countries and consisting of two regions (A and B). Both the countries and the regions are identical in all relevant characteristics. Assume that there are two options for an international regime to reduce emissions of greenhouse gases. Option 1 is a global agreement on emissions reductions and option 2 is a regime consisting of two separate agreements, one for each region.[6] We assume that, regardless of which regime is chosen, each country can choose between complying at cost c leading to a fixed reduction in emissions, and not complying.

In every period of the game, each country must choose to *cooperate* (i.e., reduce emissions) or to *defect* (not reduce emissions). Following [5], in the case of a *global agreement* where there are k participating countries and $N-k$ non-participating countries, the periodic payoff of each of the k countries playing *cooperate* is

$$dk - c\,,$$

where d is a constant ($d>0$), while each of the $N-k$ countries playing *defect* receives

$$bk\,,$$

where b is another constant ($b>0$). Similarly, with two *regional agreements*, the payoffs are

$$d\cdot(k_A+k_B)-c \text{ and } b\cdot(k_A+k_B)$$

for the k_A and k_B countries playing *cooperate* in regions A and B and the $N-k_A-k_B$ countries playing *defect*, respectively. Hence, the payoff obtained by each country if it chooses *cooperate* (*defect*) is a linear function of the total number of countries that cooperates. In an appendix we show how our analysis can be extended to the case where there are decreasing returns from cooperation. Like Barrett [5, p. 527] we assume that $d\geqslant b$. Throughout, the analysis handles the general case where $d\geqslant b$, except in Section 3.3 where the restriction $d=b$ is invoked. We also assume that $b(N-1)>dN-c>0$, meaning that full participation is *not* a Nash equilibrium of the stage game, and that full participation Pareto-dominates no participation. From the assumptions that $d\geqslant b$ and $b(N-1)>dN-c$ it follows that,

$$\text{for all } k\in\{1,\ldots,N\},\ b(k-1)>dk-c\,, \tag{1}$$

entailing that *defect* is dominant in the stage game; in particular, $c>d$. Hence, the stage game has a unique Nash equilibrium where all countries play *defect*. Furthermore, from the assumption that $dN-c>0$, it follows that this Nash equilibrium is inefficient, since all countries would gain if they all played *cooperate*. These properties entail that the stage game is a special version of the N-person Prisoners' Dilemma. We assume throughout that all countries discount future payoffs using a common discount factor δ ($0<\delta<1$).

[5]Weakly renegotiation-proofness is defined formally in Section 3.

[6]Since the countries are modeled as symmetrical players, the real question addressed by the analysis is whether two agreements are better than one. In the case of two agreements, we label them ‘‘regional’’ agreements.

We assume that compliance in the global agreement is sustained by way of the "Penance" strategy. This strategy specifies that a participating country plays *cooperate* except if another participating country has been the sole deviator from Penance in the previous period, in which case *defect* is played. Non-participating countries play *defect* after any history. Penance is closely related to the Getting Even strategy, which is used in Barrett's work.[7] It is well known that a global agreement based on Getting Even admits only a very limited number of parties (e.g., [5]).

In the case of two regional agreements, compliance is likewise sustained on the basis of a close relative to Penance. This strategy — which we refer to as "*Regional* Penance" – specifies that a participating country plays *cooperate* except if another participating country in its own region has been the sole deviator from Regional Penance in the previous period, in which case *defect* is played. As in the case of a global treaty, non-participating countries play *defect* after any history. It may be noted that if all countries use this strategy, then a deviation by a participating country in region A will be punished by other parties in region A, but not by parties to the agreement in region B, and vice versa. We show below that this is part of the equilibrium behavior.

Loosely formulated, an equilibrium is renegotiation-proof if the players as a group cannot find a better alternative [8, p. 43]. More precisely, in order to be a *weakly renegotiation-proof equilibrium* (in the sense of Farrell and Maskin [19 pp. 330–331]), a strategy profile must satisfy two requirements.[8]

(1) The strategy profile must be a subgame perfect equilibrium. Hence, every weakly renegotiation-proof equilibrium is subgame perfect (but the converse does not hold). In any repeated game with discounting it is necessary and sufficient for a strategy profile to be a subgame perfect equilibrium that no player can gain by a one-period deviation after some history. Hence, even though in the subsequent analysis we consider only the case where a deviator accepts the punishment without further deviation, we are not ruling out the possibility that a player may deviate for many periods. The point is that we need only check one-period deviations, as it follows from the theory of repeated games with discounting [1, p. 390] that a player cannot gain by a multi-period deviation if he cannot gain by some one-period deviation.
(2) Since by the first requirement the strategy profile is subgame perfect, it determines after any history a subgame perfect equilibrium — a *continuation equilibrium* — in the continuation of

[7]The Getting Even strategy that Barrett [5, p. 530] uses "... requires that country i play *cooperate* unless i has played *defect* less often than any of the other players in the past ...", and entails that a participating country is offered a future reward if it does not take part in the punishment of a deviating country. This feature does not seem essential, but leads to non-trivial calculations. The Penance strategy simplifies the Getting Even strategy by *not* letting such deviating behavior be rewarded. Likewise, specifying punishment after multilateral deviations, while certainly a real-life possibility, would in the model only complicate the analysis without changing the behavior along the equilibrium path. Like Barrett we do not seek to characterize the set of weakly renegotiation-proof equilibria. However, by analyzing the Regional Penance strategy defined below we show that the set of weakly renegotiation-proof equilibria extends beyond the kind of strategy profiles considered by Barrett [5]. A complete characterization of the set of weakly renegotiation-proof equilibria in the model is an interesting problem that we leave for future research.

[8]Various concepts of renegotiation-proof equilibrium exist in the literature (e.g., [9,19,2]; see [8] for a discussion). When we use that term in this article we always refer to the concept of weakly renegotiation-proof equilibrium, formalized for 2-person games by Farrell and Maskin [19], and extended to N-person games by Bergin and MacLeod [8, Definition 3].

the game. The second requirement is that there must not exist two continuation equilibria such that all players are better off in one continuation equilibrium than in the other. The reason is that if two such continuation equilibria existed, the players would renegotiate from one to the other. In the case of Penance and Regional Penance where all punishments last only one period, this requirement is satisfied if not all players strictly gain by choosing collectively to restart cooperation at once instead of implementing the threatened punishment when a deviation has taken place in the previous period.

In the following, we compare a global agreement to a regime based on two agreements with respect to each of these two requirements.

3.1. The subgame perfection requirement

Consider first the case of a global agreement. In this case there are two kinds of histories to check. In the first kind there is no single deviator in the previous round. Penance then prescribes that all participating countries play *cooperate*. As long as all participating countries follow this prescription, a participating country j collects a payoff of $dk - c$ in each period, where k is the number of parties in the global agreement. By contrast, suppose country j deviates from Penance in period t by playing *defect* and reverts to Penance at time $t+1$. It then gets a payoff of $b(k-1)$ in period t. In period $t+1$ it is punished by the other players and receives $d-c$. From time $t+2$ onwards cooperation is restored, meaning that country j obtains $dk-c$ in that and any future period. It is individually rational to stick to Penance unless country j strictly increases its sum of discounted payoffs in periods t and $t+1$ by deviating. Formally, this condition is satisfied if

$$(1+\delta)(dk-c) \geqslant b(k-1) + \delta(d-c).$$

Solving for k gives:

$$k \geqslant \frac{\delta d - b + c}{\delta d + d - b} = 1 + \frac{c-d}{\delta d + d - b}. \tag{2}$$

If condition (2) is fulfilled, then the strategy profile where all countries play Penance is a Nash equilibrium.

The second kind of histories to check is the one where a single deviation has taken place in period $t-1$. It follows from the definition of Penance that one of the $k-1$ participating countries that did not deviate in period $t-1$, say country i, would deviate from Penance in period t if it plays *cooperate*. However, such a deviation leads to a loss for country i in period t since it follows from (1) that $b > d2 - c$. It also leads to a loss in period $t+1$ since punishment will be triggered. Hence, it suffices to check that the country that deviated in period $t-1$ has no incentive to deviate from Penance again in period t by playing *defect*. If this country accepts the punishment and reverts to Penance at time t, then it receives $d-c$ in period t and $dk-c$ in any future period. By contrast, if it deviates for one more period before reverting to Penance, then it receives 0 at time t and $d-c$ at time $t+1$. Thus, it is individually rational for the deviating country to accept the punishment at time t, provided that

$$d - c + \delta(dk-c) \geqslant 0 + \delta(d-c).$$

Solving for k gives

$$k \geqslant \frac{\delta d - d + c}{\delta d} = 1 + \frac{c-d}{\delta d}\,. \tag{3}$$

Note that $(c-d)/\delta d \geqslant (c-d)/(\delta d + d - b)$ since we assume that $d \geqslant b$ and, by (1), $c > d$, entailing that condition (3) implies condition (2). Hence, (3) is a sufficient and necessary condition for the strategy profile where all countries play Penance to be a subgame-perfect equilibrium. This condition places a lower bound on the number of parties needed to make an agreement sustainable.[9]

Next, consider a regime with two regional agreements. Once again, there are two kinds of histories to check in order to ensure subgame perfection. In the first kind there is no single deviation in (only) one of the regions in the previous period. If a participating country j in region A deviates from Regional Penance in period t by playing *defect* and reverts to Regional Penance at time $t+1$, then in period t it collects a payoff of $b(k_A - 1 + k_B)$, where k_S is the number of parties in the agreement for region S ($S = A, B$). In period $t+1$, the deviating party pays penance by being the only country in region A to play *cooperate*, whereas cooperation continues as before in region B. Country j thus obtains a payoff of $d(1+k_B) - c$. From time $t+2$ onwards cooperation is restored, so j receives $d(k_A + k_B) - c$ in that and in any future period. Had country j not deviated, it would have received $d(k_A + k_B) - c$ in every period. Therefore, it is individually rational to stick to Regional Penance, provided that

$$(1+\delta)(d(k_A + k_B) - c) \geqslant b(k_A - 1 + k_B) + \delta(d(1+k_B) - c)\,. \tag{4}$$

We concentrate on equilibria with the same number of parties in each of the two agreements. For the analysis of weakly renegotiation-proof equilibria, this is without loss of generality, as reported in footnote 10. Setting $k_S = k_A = k_B$ and solving (4) for k_S gives:

$$k_S \geqslant \frac{\delta d - b + c}{\delta d + 2(d-b)} = 1 + \frac{c - d - (d-b)}{\delta d + 2(d-b)}\,. \tag{5}$$

Condition (5) ensures that the strategy profile where all countries use Regional Penance is a Nash equilibrium.

The second kind of histories to check is the one where a single deviation has occurred in (only) one of the two regions in the previous period. Suppose the deviation has taken place in region A in period $t-1$. It follows from the definition of Regional Penance that one of the $k_A - 1$ participating countries that did not deviate in period t, say country i, would deviate from Regional Penance in period t if it plays *cooperate*. However, such a deviation leads to a loss for country i in both period t since, by (1), $b(1+k_B) > d(2+k_B) - c$, and in period $t+1$ since punishment will be triggered. Hence, it suffices to check that the country in region A that deviated in period $t-1$ as well as all the k_B participating countries in region B have no incentive to deviate from Regional Penance in period t by playing *defect*.

Consider first the country in region A that deviated from Regional Penance in period $t-1$. If this country accepts the punishment and reverts to Regional Penance at time t, then it receives $d(1+k_B) - c$ in period t, and $d(k_A + k_B) - c$ in any future period. By contrast, if it deviates also

[9]This is one possible explanation why most multilateral treaties (including the Kyoto protocol) provide a clause stating that their entry into force requires ratification by a certain number of countries.

at time t and reverts to Regional Penance at time $t+1$, then it receives bk_B at time t and $d(1+k_B)-c$ at time $t+1$, before cooperation is restored. Hence, it is individually rational for the deviating country to accept the punishment at time t, provided that

$$d(1+k_B)-c+\delta(d(k_A+k_B)-c)\geqslant bk_B+\delta(d(1+k_B)-c)\,. \tag{6}$$

Setting $k_S=k_A=k_B$ and solving for k_S gives

$$k_S\geqslant\frac{\delta d-d+c}{\delta d+d-b}=1+\frac{c-d-(d-b)}{\delta d+d-b}\,. \tag{7}$$

Turn now to the k_B participating countries in region B, given that a deviation in region A has occurred in period $t-1$. If all countries use Regional Penance from period t, then a participating country j in region B receives $d(1+k_B)-c$ in period t, and $d(k_A+k_B)-c$ in period $t+1$. By contrast, if country j deviates from Regional Penance in period t by playing *defect* and reverts to Regional Penance at time $t+1$, then it receives $b(1+k_B-1)$ at time t and $d(k_A+1)-c$ at time $t+1$. In either case, cooperation is reestablished in both regions in period $t+2$. Thus, country j has no incentive to deviate from Regional Penance if

$$d(1+k_B)-c+\delta(d(k_A+k_B)-c)\geqslant bk_B+\delta(d(k_A+1)-c)\,.$$

Setting $k_S=k_A=k_B$ and solving for k_S yields (7).

Note that $(c-d-(d-b))/(\delta d+d-b)\geqslant(c-d-(d-b))/(\delta d+2(d-b))$ since we assume that $d\geqslant b$ and, by (1), $b>d2-c$, entailing that condition (7) implies condition (5). Hence, (7) is a sufficient and necessary condition for the strategy profile where all countries play Regional Penance to be a subgame-perfect equilibrium. This condition places a lower bound on the number of parties in a regime based on two agreements.

3.2. *The renegotiation-proofness requirement*

The requirement of renegotiation-proofness is satisfied in the present setting if not all players strictly gain by choosing collectively to restart cooperation at once instead of implementing the threatened punishment when a deviation has taken place in the previous period. Hence, at least one country must be at least as well off with punishment as with renegotiation. It is *not* necessary that all countries but the one being punished are at least as well off with punishment as with renegotiation.

Consider the global agreement, and assume that a participating country j deviates in period $t-1$. Punishment implies that all participating countries but the deviator switch from playing *cooperate* to playing *defect*. This implies that not only the deviator, but all non-participating countries are worse off with punishment than with renegotiation. Weak renegotiation-proofness thus requires that the punishing participating countries are at least as well off with punishment as with renegotiation, because otherwise all countries would gain by restarting cooperation at once. If the punishment is carried out in accordance with Penance, then each of the punishing participating countries receives a payoff of b in period t. By contrast, if cooperation is restarted immediately, then each of them receives $dk-c$ in period t. In either case, each participating country receives $dk-c$ in every future period, so only period t needs to be considered. Therefore, weak renegotiation-proofness requires that

$$b\geqslant dk-c\,,$$

which can be restated as

$$k \leqslant \frac{b+c}{d} \,. \tag{8}$$

Thus, weak renegotiation proofness requires that the number of participating countries not be too large. The intuition behind this result is that, other things being equal, a larger number of participating countries increases the payoff associated with cooperation. A large number of participating countries thus makes renegotiation attractive, thereby undermining the (collective) incentive to punish a deviation.

Next, consider regional agreements. Assume that country j is a party to the agreement in region A, and that this country has defected in period $t-1$. If the punishment is carried out in accordance with Regional Penance, then each of the punishing participating countries in region A receives $b(1+k_B)$ in period t. By contrast, if cooperation is restarted at once, each of them receives $d(k_A+k_B)-c$ in period t. Thus, weak renegotiation-proofness requires that

$$b(1+k_B) \geqslant d(k_A+k_B)-c \,.$$

With the same participation in both regions, this condition reduces to

$$k_S \leqslant \frac{b+c}{2d-b} \,. \tag{9}$$

3.3. Summary of main results with numerical illustrations

While if follows from conditions (3) and (7) that the subgame perfection requirement places a lower bound on the number of participating countries, by conditions (8) and (9) the renegotiation-proofness requirement places an upper bound on this number. Combining the subgame perfection and renegotiation proofness requirements, we find that a global treaty modeled by the Penance strategy profile must satisfy

$$\frac{\delta d - d + c}{\delta d} \leqslant k \leqslant \frac{b+c}{d} \,. \tag{10}$$

It follows from condition (10) that the strategy profile where all countries play Penance is a weakly renegotiation-proof equilibrium for some δ close (but not equal) to 1 if and only if

$$\frac{c}{d} < k \leqslant \frac{b}{d} + \frac{c}{d} \,. \tag{11}$$

Because $d \geqslant b$, there is at most one integer k^* satisfying (11). Hence, whenever a global agreement exists, the number of parties is always fully determined. Moreover, as the stage game is a Prisoners' Dilemma, so that $dN - c > 0$, there exists a unique integer $k^* \leqslant N$ satisfying (11) if $d = b$ (since then the right hand side of (11) becomes $1 + c/d$). This establishes that a global treaty can always be implemented as a weakly renegotiation-proof equilibrium provided that $d = b$ and the countries are sufficiently patient.

Similarly, two symmetrical regional agreements modeled by Regional Penance must satisfy

$$\frac{\delta d - d + c}{\delta d + d - b} \leqslant k_S \leqslant \frac{b+c}{2d-b} \,. \tag{12}$$

Condition (12) implies that the strategy profile where all countries play Regional Penance is a weakly renegotiation-proof equilibrium for some δ close (but not equal) to 1 if and only if

$$\frac{c}{2d-b} < k_S \leqslant \frac{b}{2d-b} + \frac{c}{2d-b}\,. \tag{13}$$

Because $d \geqslant b$, there is at most one integer k_S^* satisfying (13). Hence, whenever a regime based on two symmetrical regional agreements exists, the number of parties is always fully determined. Moreover, if $d = b$ and $N \geqslant 2(1 + c/d)$, then there exists a unique integer k_S^* satisfying (13) and $k_S^* + k_S^* \leqslant N$ (since with $d = b$ (13) becomes $c/d < k_S \leqslant 1 + c/d$). Hence, with enough countries, two symmetrical agreements can always be implemented as a weakly renegotiation-proof equilibrium provided that $d = b$ and the countries are sufficiently patient.[10]

We thus see in both regimes that the renegotiation-proofness requirement in combination with the subgame perfection requirement determines the number of parties, thus leaving a very narrow set of possibilities. If $k^* < N$ in the case of a global treaty, or if $k_S^* + k_S^* < N$ in the case of two symmetrical agreements, then there exist multiple equilibria: even though condition (11) determines the *number* of participating countries in a global agreement, and condition (13) the *number* of participating countries in the two regional agreements, neither condition determines *which* countries participate.

It follows from conditions (10) and (12) that, for $d = b$, the lower and upper bounds on each regional agreement are identical to the corresponding bounds on the global agreement. In this case the number of participating countries required and admitted by the regional agreements regime is twice that of a global agreement. To illustrate, assume that $\delta = 0.95$, $c = 8$, and $d = b = 1$. We then find that a global regime requires at least nine participating countries. At the same time, it admits a maximum of nine countries as well. With two agreements, by contrast, the minimum and maximum number of parties is nine in each region, meaning that the total number of participating countries is 18.

If $d > b$, then the difference between the two types of regime can be smaller. For example, if $\delta = 0.95$, $c = 8$, $d = 1.25$, and $b = 1$, then the minimum and maximum number of parties with regional agreements becomes 12 (six in each region). By contrast, with a global agreement the minimum and maximum number of parties is seven.

[10]If we had done the analysis *without* setting $k_S = k_A = k_B$, then, instead of (13), the following inequalities would have been necessary conditions:

$$\frac{c-(d-b)k_B}{d} < k_A \leqslant \frac{b}{d} + \frac{c-(d-b)k_B}{d}\,, \tag{14}$$

$$\frac{c-(d-b)k_A}{d} < k_B \leqslant \frac{b}{d} + \frac{c-(d-b)k_A}{d}\,. \tag{15}$$

However, the right inequality of (14) contradicts the left inequality of (15) if $k_A > k_B$ (and vice versa if $k_A < k_B$). It follows that two *asymmetric* agreements cannot be implemented as a weakly renegotiation-proof equilibrium. This means that we can set $k_S = k_A = k_B$ without loss of generality.

3.4. Regional agreements and Pareto dominance

The model analyzed in this paper suggests that a regime based on regional agreements can be beneficial to all countries. A global agreement with k^* participating countries is (strongly) Pareto dominated by a situation where participation is extended to n countries (where $k^* < n \leqslant N$) if this makes every country better off. It follows immediately that the k^* countries that play *cooperate* both before and after such an enlargement gain by increased participation. The same is true for the $N - n$ countries that play *defect* both before and after the enlargement. The reason is simply that, from the point of view of these countries, the enlargement increases the number of *other* countries that play *cooperate*. The question is therefore whether the $n - k^*$ countries that switch from *defect* (and receiving a periodic payoff of bk^*) to *cooperate* (and obtaining $dn - c$) gain by doing so. Hence, a global treaty with k^* participating countries is Pareto dominated by an enlargement to n members if

$$dn - c > bk^* .$$

Below we compare the following two situations,

- a global regime where k^* countries cooperate,
- a regional regime where $k_A^* > 1$ countries form a new agreement in region A, and the $k_B^* = k^*$ members of the global agreement constitute another agreement in region B,

and show that the former is always Pareto dominated by the latter.

As argued above, the k_B^* countries in region B included as parties under both regimes, as well as the $N - k_A^* - k_B^*$ countries that are non-parties in both cases, prefer the regional regime. Hence, the real test is whether the regional regime is acceptable to the k_A^* members of the additional agreement, i.e., to countries that are allowed to free ride in the first situation, but required to cooperate in the second. With a global agreement, each of these countries receives a periodic payoff of bk^*. With two regional agreements, each of them obtains $d(k_A^* + k_B^*) - c$. Hence, if

$$d(k_A^* + k_B^*) - c > bk^* , \tag{16}$$

then each of these countries is better off being a party to the new regional agreement than as a free rider on a global agreement. It follows that, if condition (16) is fulfilled, a regime based on regional agreements Pareto dominates a global agreement. We show below, through a verbal argument and an analytical demonstration, that this condition is always fulfilled.

For a verbal argument, recall that it is feasible for any participating country j in the additional agreement in region A to choose a perpetual deviation by always playing *defect*. If country j starts playing *defect* in period t, then all other countries in region A will also play *defect* from period $t + 1$, thereby effectively terminating the additional agreement. Such a perpetual deviation by country j from the agreement in region A would give country j a higher payoff than it would obtain with the global agreement, since it benefits from the cooperation of the $k_A^* - 1$ other countries in region A in period t, and from the cooperation of the $k_B^* = k^*$ participating countries in region B in all periods. However, it follows from the subgame perfection requirement that such a deviation is not profitable, implying that a perpetual play of *cooperate* in the new agreement in region A is even more advantageous. Hence, whenever the subgame perfection requirement is fulfilled, a single global agreement is Pareto dominated by a regional regime where a new regional agreement is introduced.

For an analytical demonstration, multiplying (4) by $1-\delta$ and rearranging terms gives

$$d(k_A^* + k_B^*) - c \geqslant (1-\delta)b(k_A^* - 1 + k_B^*) + (1-\delta)\delta\big(d(1+k_B^*) - c\big) + \delta^2\big(d(k_A^* + k_B^*) - c\big). \tag{17}$$

By rearranging (6), we get

$$(1-\delta)\big(d(1+k_B^*) - c\big) + \delta\big(d(k_A^* + k_B^*) - c\big) \geqslant bk_B^*. \tag{18}$$

Combining inequalities (17) and (18) yields

$$d(k_A^* + k_B^*) - c \geqslant (1-\delta)b(k_A^* - 1 + k_B^*) + \delta b k_B^*,$$

which can be rewritten as

$$d(k_A^* + k_B^*) - c \geqslant bk_B^* + (1-\delta)b(k_A^* - 1).$$

Condition (16) follows with *strict* inequality since $k_B^* = k^*$, $k_A^* > 1$, and $\delta < 1$. Hence, even the participating countries in region A *strictly* gain by forming the additional agreement.

4. Summary and discussion

This paper has used a simple dynamic game-theoretic model to analyze how a regime based on two agreements would affect the prospects for international environmental cooperation. A main finding is that two treaties are able to encompass a larger number of parties than a single global treaty, assuming that the cost of cooperation is the same with both types of regime, and that all countries are identical in all relevant characteristics. We have also shown that a system with two agreements can Pareto dominate a regime based on a global treaty. These results support the findings of previous research by Carraro and others (e.g., [11,12,17]) using a reduced-stage game framework.

We end this paper by commenting briefly on three issues that relate to the analysis in Section 3. First, while the model analyzed in this paper includes only two regions, one might reasonably ask why one should stop there. Why not design a regime with an even larger number of agreements? This idea certainly cannot be dismissed on the basis of the above model. We have seen, however, that each regional agreement requires a certain minimum number of parties. Hence, with a fixed number of countries to select from there is an upper limit to the number of agreements that can be sustained in equilibrium.

Second, an important driving force in Barrett's [5] model of a global treaty, as well as in our model of regional cooperation, is the way in which the agreement(s) are enforced. The maximum number of signatories is very limited in Barrett's model because a single violation triggers punishment by *all* other parties. This drives the number of participating countries down via the renegotiation-proofness requirement. The reason is that for each participating country the payoff obtained when all participating countries play *cooperate* must not exceed the payoff of playing *defect* when only the country being punished plays *cooperate*. Because $d>0$, this requirement can only be satisfied if (very) few countries participate. If *not all* participating countries punish a deviator, then more countries will cooperate in the punishment phase. This means that, in contrast to Barrett's model, the renegotiation proofness requirement becomes less strict and admits a larger number of participating countries. Regional cooperation means that only some of the participating countries punish a deviator: even if a violation in region A is punished by all parties to the agreement for region A, cooperation in region B continues undisturbed. It is this feature that allows the number of participating countries to become larger with a regime based on regional agreements.

Finally, in addition to the regional agreements approach highlighted in this paper, there are also other avenues for broadening the number of parties that can be sustained in weakly renegotiation-proof equilibrium. For example, the maximum number of parties in a global agreement might be increased by designing an enforcement scheme whereby a single defection triggers punishment only by a subset of the other parties.[11] Indeed, even a global agreement could — in principle — be enforced regionally, so the effects analyzed here could also be achieved within the setting of a global agreement. However, it seems rather implausible that a system of regional enforcement would be adopted except as part of a regime based on regional agreements. Moreover, informal arguments suggest that a regime based on regional agreements is likely to have other advantages as well. Countries that are in close geographic proximity also tend to be culturally close, have similar economic and political systems, and therefore have similar preferences.[12] These factors might reduce uncertainty, enhance trust, and reduce negotiation costs. Countries may thus not only be more likely to comply with a regional agreement, but also more inclined to join a regional agreement in the first place. This further strengthens the conclusions of the formal analysis in Section 3, where there are no cultural, economic, or political ties between any subsets of countries. Thus, regional cooperation might be a good alternative — or supplement — to global IEAs like the Kyoto Protocol.

Acknowledgments

We are indebted to two anonymous referees of this journal for detailed comments on previous drafts. We are also grateful to Scott Barrett, Leif Helland, Aanund Hylland, Ottar Mæstad, Kevin Siqueira, Jon Vislie, and colleagues at CICERO, in particular Odd Godal and Asbjørn Torvanger, for helpful discussions and suggestions. Finally, we thank the Research Council of Norway for financial support, Lynn Nygaard for editorial assistance, and Frode Rørvik for computer support.

[11]In fact, in the appendix to [5] this type of regime is examined, but rejected on the grounds that it fails to be strongly collectively rational.

[12]Of course, this is not always the case. In the Middle East, for instance, some neighboring countries obviously have very different preferences on certain issues.

Appendix A. Decreasing returns from cooperation

Throughout this paper, it has been assumed that the utility obtained by each country if it chooses to *cooperate* (*defect*) is a linear function of the total number of countries that cooperate. What happens if there are decreasing returns from cooperation?

Let $f(k)$ denote the gain accruing to each country if k is the total number of countries cooperating, when compared to the situation where no country cooperates. Hence, in a global agreement, the payoff of each of the k countries playing *cooperate* is

$$f(k) - c\,,$$

while each of the $N - k$ countries playing *defect* receives

$$f(k)\,.$$

Likewise, with two regional agreements, the payoffs are

$$f(k_A + k_B) - c \text{ and } f(k_A + k_B)$$

for the k_A and k_B countries playing *cooperate* in regions A and B and the $N - k_A - k_B$ countries playing *defect*, respectively. This means that the setting of the main text is generalized by not imposing a linear relationship between the gains from cooperation and the number of cooperating countries. On the other hand, the gain accruing to a country depends now only on the total number of cooperating countries, not also on whether the country itself plays *cooperate* (as in the main text when $d > b$).

The function $f(\cdot)$ is assumed to satisfy

$$f(0) = 0 \text{ and for all } k \geqslant 1,\ 0 < f(k+1) - f(k) \leqslant f(k) - f(k-1)\,, \tag{A.1}$$

meaning that the gain from an additional country cooperating is a non increasing function of the number of countries cooperating. We refer to this as non-increasing returns from cooperation. If $d = b$, then the analysis of the main text is the special case where $f(k) = dk = bk$. Moreover, consider the following assumptions:

$$f(1) < c\,, \tag{A.2}$$

$$f(N) - f(1) > c\,, \tag{A.3}$$

$$\text{There exists an even integer } N^* \leqslant N \text{ s.t. } f(N^*) - f(1 + N^*/2) > c\,. \tag{A.4}$$

It is straightforward to check that (A.1)–(A.3) ensure that the stage game is a Prisoners' Dilemma, while (A.1) and (A.4) imply (A.3).

Assume (A.1)–(A.3). Consider a global treaty based on Penance with k participating countries. It follows from the corresponding analysis in the main text that subgame perfection requires that

$$f(k-1) - f(k) + c \leqslant \delta(f(k) - f(1)) \text{ and } -f(1) + c \leqslant \delta(f(k) - f(1))\,,$$

where, by (A.1), the first inequality implies the second one. Hence, if

$$2f(k) - f(k-1) > f(1) + c\,, \tag{A.5}$$

there exists $\delta < 1$ such that the strategy profile where all countries play Penance is a subgame perfect equilibrium. On the other hand, for this strategy profile to be a weakly renegotiation-proof equilibrium, it must also hold that $f(1) \geqslant f(k) - c$, or equivalently,

$$f(k) \leqslant f(1) + c\,. \tag{A.6}$$

To see that there exists some k^* satisfying both (A.5) and (A.6), note that (A.3) implies that $f(k+1) \leqslant f(1) + c$ for $k = 0$ and $f(k+1) > f(1) + c$ for $k = N - 1$. Let k^* be the smallest integer satisfying $f(k^* + 1) > f(1) + c$, entailing that (A.6) is satisfied for $k = k^*$. It follows that (A.5) is also satisfied for $k = k^*$ since, by (A.1),

$$2f(k^*) - f(k^* - 1) \geqslant f(k^* + 1)\,.$$

Furthermore, since $f(k)$ is strictly increasing, we have that k^* determined in this way is the largest k for which (A.5) and (A.6) are satisfied.

Assume now the stronger set of assumptions (A.1), (A.2), and (A.4). In Regional Penance with k_S participating countries in each agreement, it follows from the corresponding analysis in the main text that subgame perfection requires that

$$f(2k_S - 1) - f(2k_S) + c \leqslant \delta(f(2k_S) - f(1 + k_S)) \text{ and}$$
$$f(k_S) - f(k_S + 1) + c \leqslant \delta(f(2k_S) - f(1 + k_S))\,,$$

where, by (A.1), the first inequality implies the second one. Hence, if

$$2f(2k_S) - f(2k_S - 1) > f(1 + k_S) + c\,, \tag{A.7}$$

there exists $\delta < 1$ such that the strategy profile where all countries play Regional Penance is a subgame perfect equilibrium. On the other hand, for this strategy profile to be a weakly renegotiation-proof equilibrium, it must also hold that $f(1 + k_S) \geqslant f(2k_S) - c$, or equivalently,

$$f(2k_S) \leqslant f(1 + k_S) + c\,. \tag{A.8}$$

To see that there exists some k_S^* satisfying both (A.7) and (A.8), note that (A.1) and (A.4) imply that, if $0 < k_B \leqslant N^*/2$, then $f(k_A + 1 + k_B) \leqslant f(1 + k_B) + c$ for $k_A = 0$ and $f(k_A + 1 + k_B) > f(1 + k_B) + c$ for $k_A = N^*/2 - 1$. Let $\kappa(k_B)$ be the smallest integer satisfying $f(\kappa(k_B) + 1 + k_B) > f(1 + k_B) + c$, entailing that

$$f(\kappa(k_B) + k_B) \leqslant f(1 + k_B) + c\,. \tag{A.9}$$

It follows that

$$2f(\kappa(k_B) + k_B) - f(\kappa(k_B) - 1 + k_B) > f(1 + k_B) + c \tag{A.10}$$

also holds since $2f(\kappa(k_B) + k_B) - f(\kappa(k_B) - 1 + k_B) \geqslant f(\kappa(k_B) + 1 + k_B)$ by (A.1). By (A.1) and the definition of $\kappa(\cdot)$ we have that

$$0 < \kappa(k_B) \leqslant \kappa(k_B + 1) < N^*/2 \quad \text{if } 0 \leqslant k_B \leqslant N^*/2 - 1\,. \tag{A.11}$$

Consider the sequence $\{k^0, k^1, k^2, \ldots\}$ defined by $k^0 = 0$ and, for all $n \geqslant 1$, $k^n = \kappa(k^{n-1})$. By (A.10), $\{k^0, k^1, k^2, \ldots\}$ is a non-decreasing sequence that converges to k_S^* satisfying

$$0 < k_S^* = \kappa(k_S^*) < N^*/2\,. \tag{A.12}$$

Now, (A.9), (A.10), and (A.12) imply that (A.7) and (A.8) are satisfied for $k_S = k_S^*$.

To compare the maximum number of participating countries in a single global agreement vs. two regional agreements, assume (A.1), (A.2), and (A.4) (which, as noted above, imply (A.3)). It follows from the definition of $\kappa(\cdot)$ that the maximum number of participating countries under a global agreement is given by $k^* = \kappa(0)$, while the regional agreements regime admits $k_S^* = \kappa(k_S^*)$ participating countries in each region. Since, by (A.11), $k^* = \kappa(0) \leqslant k_S^* = \kappa(k_S^*)$, we reach the conclusion that *two regional agreements allow for at least twice as many participating countries as compared to a single global agreement*, when there are non-increasing gains from cooperation.

The intuition for this finding is as follows. With decreasing gains from cooperation, the temptation for a given number of punishing countries to restart cooperation at once after a deviation has occurred, is smaller if cooperation continues unaffected in a separate regional agreement. Therefore, participation in *each* of two regional agreements can be *strictly larger* than the maximal participation in a single global agreement.

A numerical example can be used to illustrate this possibility. It can be checked that (A.1), (A.2), and (A.4) are satisfied with $c = 1.01$, $f(k) = \sqrt{k}$, and $N = 18$. Then $k = 4$ is the largest number of participating countries satisfying (A.5) and (A.6), while the sequence $\{k^0, k^1, k^2, \ldots\}$, defined by $k^0 = 0$ and, for all $n \geqslant 1$, $k^n = \kappa(k^n - 1)$, equals $\{0, 4, 6, 7, 7, \ldots\}$. Hence, (A.7) and (A.8) are satisfied with $k_S = 7$ countries in each of the two regional agreements. Actually, $k_S = 8$ is the largest number of participating countries satisfying (A.7) and (A.8); hence, our algorithm in the case of Regional Penance does not necessarily converge to the maximum number of participating countries, but suffices to show that maximum total participation at least doubles when moving from a single global treaty to two regional agreements.

References

[1] D. Abreu, On the theory of infinitely repeated games with discounting, Econometrica 56 (1988) 383–396.
[2] G.B. Asheim, Extending renegotiation-proofness to infinite horizon games, Games Econ. Behav. 3 (1991) 278–294.
[3] S. Barrett, Self-enforcing international environmental agreements, Oxford Econ. Pap. 46 (1994) 878–894.
[5] S. Barrett, A theory of full international cooperation, J. Theoret. Polit. 11 (1999) 519–541.
[6] S. Barrett, Consensus treaties, J. Institut. Theoret. Econ. 158 (2002) 529–547.
[7] S. Barrett, Environment & Statecraft: The Strategy of Environmental Treaty-Making, Oxford University Press, New York, 2003.
[8] J. Bergin, B. MacLeod, Efficiency and renegotiation in repeated games, J. Econ. Theory 61 (1993) 42–73.
[9] B.D. Bernheim, D. Ray, Collective dynamic consistency in repeated games, Games Econ. Behav. 1 (1989) 295–327.
[10] F. Bloch, Non-cooperative models of coalition formation in games with spillovers, in: C. Carraro, D. Siniscalco, (Eds.), New Directions in the Economic Theory of the Environment, Cambridge University Press, Cambridge, UK, 1997.
[11] C. Carraro, The structure of international agreements on climate change, in: C. Carraro, (Ed.), International Environmental Agreements on Climate Change, Kluwer Academic Publishers, Dordrecht 1999.
[12] C. Carraro, The economics of coalition formation, in: J. Gupta, M. Grubb, (Eds.), Climate Change and European Leadership, Kluwer Academic Publishers 2000.
[13] C. Carraro, D. Siniscalco, The international dimension of the environmental policy, European Econ. Rev. 26 (1992) 379–387.
[14] C. Carraro, D. Siniscalco, Strategies for the international protection of the environment, J. Public Econ. 52 (1993) 309–328.
[17] C. Carraro, D. Siniscalco, International environmental agreements: Incentives and political economy, Europ. Econ. Rev. 42 (1998) 561–572.
[18] P. Chander, H. Tulkens, Theoretical foundations of negotiations and cost sharing in transfrontier pollution problems, Europ. Econ. Rev. 36 (1992) 388–398.
[19] J. Farrell, E. Maskin, Renegotiation in repeated games, Games Econ. Behav. 1 (1989) 327–60.
[20] M. Finus, Game Theory and International Environmental Cooperation, E. Elgar, Cheltenham 2001.

[21] M. Finus, B. Rundshagen, Renegotiation-proof equilibria in a global emission game when players are impatient, Environ. Resource Econ. 12 (1998) 275–306.
[22] M. Finus, B. Rundshagen, Endogenous coalition formation in global pollution control: A partition function approach, in: C. Carraro, (Ed.), Endogenous Formation of Economic Coalitions, E. Elgar, Cheltenham, 2003.
[23] M. Hoel, International environmental conventions: the case of uniform reductions of emissions, Environ. Resource Econ. 2 (1992) 141–159.
[24] H. Tulkens, An economic model of international negotiations relating to transfrontier pollution, in: K. Krippendorff, (Ed.), Communication and Control in Society, Gordon & Breach, New York, 1979.