Center for
Mathematical Economics
Working Papers

545

September 2015

Local and Global Pollution and International Environmental Agreements in a Network Approach

Michael Günther and Tim Hellmann

Center for Mathematical Economics (IMW)
Bielefeld University
Universitätsstraße 25
D-33615 Bielefeld · Germany

e-mail: imw@uni-bielefeld.de
http://www.imw.uni-bielefeld.de/wp/
ISSN: 0931-6558

Local and Global Pollution and International Environmental Agreements in a Network Approach

Michael Günther[*] Tim Hellmann[†]

September 3, 2015

Abstract

Increasing concerns about climate change have given rise to the formation of International Environmental Agreements (IEAs) as a possible solution to limit global pollution effects. In this paper, we study the stability of IEAs in a repeated game framework where we restrict to strategies which are simple and invariant to renegotiation. Our main contribution to the literature on IEAs is that we allow for heterogeneous patterns of pollution such that additional to a global effect of pollution there are local pollution effects represented by a network structure. We show that stable IEAs exist if the network structure is balanced. Too large asymmetries in the degree of local spillovers may however lead to non-existence of stable structures. The generality of our approach allows for several applications to general problems in the provision of public goods.

Keywords: International environmental agreements; Weak renegotiation-proofness; Networks; Coalition structures

JEL classification: C72, C73, D85, F53, Q54

[*] Center for Mathematical Economics (IMW) and Bielefeld Graduate School of Economics and Management (BiGSEM), Bielefeld University, michael.guenther@uni-bielefeld.de. This research was carried out within the International Research Training Group "Economic Behavior and Interaction Models" (EBIM) financed by the German Research Foundation (DFG) under contract GRK 1134/2.

[†] Center for Mathematical Economics (IMW), Bielefeld University, tim.hellmann@uni-bielefeld.de. The authors would like to thank Florian Gauer, Christoph Kuzmics and Jakob Landwehr, as well as the participants of various seminars at Bielefeld University and University Paris 1 Panthéon-Sorbonne for valuable comments and suggestions.

1 Introduction

Rising concerns about climate change has led politicians worldwide to rethink their countries' emission of greenhouse gases and air pollution. Doing what is best for their own countries' interest, however, does not fully internalize the global effects of the emissions and hence their optimal policy will not reduce pollution efficiently. In other words, countries free-ride on others' abatement efforts, similar to the case of private provision of public goods. To overcome this dilemma and achieve more efficient pollution abatement, several International Environmental Agreements (IEAs) have been proposed and formed in recent years.[1]

Besides their global effects, many forms of pollution have additional negative effects on countries within the same region of the polluting source. Air pollution, for instance, can cause smog, acid deposition and eutrophication which are mostly experienced locally while the global effects (e.g. global warming) are endured worldwide. Short-lived climate pollutants such as black carbon, methane and tropospheric ozone have both a local and global impact. Their effects on "regional and global climate, through both direct interaction with atmospheric radiation and indirect effects related to changes in cloud properties are a growing concern" (Committee on the Significance of International Transport of Air Pollutants; National Research Council, 2009). As another example, a nuclear power plant causes higher negative effects in nearby regions by danger of malfunctioning compared to the global risk. The presence of these local spillovers hence plays a non-negligible role and adds additional heterogeneity to the problem of forming IEAs.

In this paper, we ask which IEAs form by purely self-interested countries when the negative externalities of pollution have a local and a global component. In our model, an IEA coordinates the abatement efforts of its members to maximize joint utility. We use a repeated game approach of abatement efforts to study the stability of these IEAs. By stability we mean that an IEA shall be self-enforcing, i.e. no member shall have an incentive to deviate from cooperation and renegotiation shall be prevented. Formally, an IEA is stable, if it can be supported by a weak renegotiation-proof equilibrium in the repeated game, where we focus on simple strategies, i.e. on one-period punishment paths.

Which countries are affected by the local externality of pollution is represented by a network: a link between two countries indicates whether these countries' pollution affects each other locally. Here, a link could mean that two countries share the same border or are within some distance which is critical for the local externality. Given a local spillover structure, we derive optimal punishment strategies such that the grand coalition of all countries can be supported by a subgame perfect equilibrium. In contrast to the general literature without a local spillover structure, global cooperation may fail to be a weakly renegotiation-proof equilibrium in very asymmetric networks. However, we also show

[1] Examples include the Oslo Protocol on sulfur reduction in Europe (also including other states) in 1994, the Montreal Protocol on the depletion of the ozone layer in 1987 and the Kyoto-Protocol on the reduction of greenhouse gases in 1997.

that it can always be sustained in regular networks.

The additional local spillover structure adds heterogeneity to the problem of formation of IEAs which can be shown to have interesting effects such that in some asymmetric structures, the global IEA is not sustainable as an equilibrium. As global pollution can be seen as a perfect public bad, the local side of it has the characteristics of a local public bad. Since reducing pollution has the characteristic of a public good, we also contribute to the problem of public good provision when the public good has both a local and a global component. To our knowledge, including both aspects in one model is also new to the literature of public goods.

Our results have important policy implications. When contemplating an IEA, strict rules have to be imposed in order to prevent deviation. These rules must specify the consequences of deviating from the agreed reductions and shall make use of the local spillover effects. With respect to welfare, we show in Section 6 that it is indeed better to first appoint neighbors for punishment of a deviation before non-neighbors shall punish. Moreover, these punishment strategies can also be invoked in order to convince other countries to join the IEA and to use the cooperation strategy. The process of how such coalitions may emerge and grow from a local to a global level can also be modeled in our network approach. We discuss this in our extensions in Section 7.

More generally, the results may serve as a benchmark that can be useful in future analyses of IEAs. It may very well be extended in several ways that we discuss in our Conclusion (Section 8). Moreover, it can easily be transferred to other problems of public good provision and may support a better understanding of free-riding problems.

The rest of the paper is organized as follows: first, we further elaborate on the issue of local and global pollution and discuss related literature as well as our contributions. In Section 3 we introduce the basic model of a single-stage game. In Section 4 we extend the model to an infinitely repeated game and derive conditions on existence of weakly renegotiation-proof equilibria for several prominent networks. Section 5 focuses on the welfare-maximizing global IEA. In Section 6 we analyze welfare implications of different network structures and in Section 7 we discuss several extension possibilities. Finally, Section 8 concludes. All proofs are presented in the Appendix.

2 Background and Literature Review

International Environmental Agreements (IEAs) have been analyzed in various game-theoretic models over the past two decades. Starting with the seminal paper by Barrett (1994), several authors have studied the free-rider problem when joining an agreement by studying both one-shot and repeated games. For a good overview of the game-theoretic literature on environmental economics we refer to recent literature surveys such as for example Jørgensen et al. (2010) or Benchekroun and Long (2012).

A majority of the models in the literature tackles the problem of air pollution, caused by the emission of greenhouse gases from fossil fuel combustion. While some models have at least abstracted from the stark assumption of homogenous countries and introduced asymmetries to account for different impact and contribution levels of pollution (e.g., McGinty, 2007; Hannesson, 2010), the implications of geographical distance to the sources of air pollution have not been largely accounted for.

However, there is broad scientific evidence for the importance of regional characteristics for several air pollution effects. Most importantly, short-lived air pollutants, that include methane, black carbon and tropospheric ozone, have a significant local or regional impact besides contributing to global problems such as climate change (see e.g., Kühn et al., 2013, for a study of emissions on local and global aerosol properties for China and India). Other examples for the regional effects of air pollution include the ozone level. For instance, the ozone level of the Mediterranean region is not only affected by local emissions but also perturbed by long-range pollution import from Northern Europe, North America and Asia (Richards et al., 2013).

Summarizing the above evidence we can conclude that the consideration of local spillover effects in addition to global externalities of emissions is crucial to better understand and represent the incentives to form IEAs. While Yang (2006) considers an optimal control problem where countries provide negatively (!) correlated local and global stock externalities (his example is CO_2 and SO_2), Dockner and Nishimura (1999) consider a dynamic game model where each country contributes to a domestic stock of pollution. Both, however, do not consider the possibility of forming an IEA to reduce pollution.

Hence, to our knowledge there exists no game theoretic model that incorporates both a local and global spillover effect of air pollution in a standard coalition formation game for an IEA. This however seems to be crucial in understanding possible solutions to the problem of reducing pollution as for example Bollen et al. (2009) show in a cost-benefit analysis, concluding that "combined climate and local air pollution policy generates extra benefits in terms of climate change mitigation." They therefore recommend that policies need to be designed such that they jointly implement both global climate change and local air pollution strategies.

Considering only global pollution as a repeated game, several works have studied IEAs as a coalition which may punish possible deviators by returning to pollution strategies. Often, a grim-trigger-strategy is considered such that all members of an IEA punish a deviator before returning to cooperation. This has been found to limit outcomes in terms of cooperating countries in equilibrium (e.g., Barrett, 1994, 1999), as the more countries punish a deviator, the fewer countries cooperate in the punishment phase which then lowers the punishing countries' payoffs in this phase. To lower the incentives for renegotiation, several authors studied different punishment strategies where not all signatories punish a deviator. Among those are Asheim and Holtsmark (2009) and Froyn and Hovi (2008). Another example is Asheim et al. (2006), where artificially two regions are introduced in order to restrict punishment to be executed only by a subset of IEA members.

By incorporating the regional effects in our model, however, it comes very natural to use the regional structure for punishment patterns.

The application of network theory to problems of public goods is not new to the literature. Several authors analyze the provision of public goods in a network and study a local spillover effect where players can only benefit from their direct neighbors' provisions (e.g., Allouch, 2015; Bramoullé and Kranton, 2007; Bloch and Zenginobuz, 2007; Elliott and Golub, 2013). However, none of these include a global spillover effect that would be necessary for an adequate representation of the pollution problem. We therefore contribute to the climate change literature by incorporating elements of the network theory, an issue that is becoming more and more interesting to researchers of that field (see Currarini et al., 2014).

3 A Pollution Game of Local and Global Spillovers

3.1 Model Setup

We consider an economy with a finite set of countries N, which are denoted by $i = 1, \ldots, n$. Countries are heterogeneous with respect to their size (i.e. satiation level of consumption) and their position in the local spillover network. We assume that countries are represented by one individual.[2] Each country derives benefits from consuming a good $x_i \in \mathbb{R}_+$ with marginal benefits assumed to be decreasing. Likewise, we assume decreasing returns from additional abatement efforts (i.e. consumption reduction). Benefits of consumption are therefore represented by the quadratic and concave function

$$B_i(x_i) = -\frac{1}{2}\left(\bar{x}_i - x_i\right)^2,$$

where $\bar{x}_i \in \mathbb{R}$ is an exogenously fixed satiation level which represents the first-best emission level if there would be no pollution effects of consumption – or at least there would be no concern for them.

Note. In the following we will make use of the following notation: $x \in \mathbb{R}^n = (x_1, \ldots, x_n)$ describes the output vector of all countries. For a subset $A = \{i_1, \ldots, i_l\} \subseteq N$, the vector $x_A = \left(x_{i_1}, \ldots, x_{i_l}\right)$ is the output vector of all countries in A. Also, we use the following abbreviation for the output vector of all countries but country i: $x_{-i} = x_{N\setminus\{i\}}$.

Consuming x_i emits air pollutants and thus contributes to the stock of pollution which is accumulated on a local and global level.[3] While benefits from consuming x_i are private, the emission of pollutants has spillover effects on all other countries. All countries equally suffer from the global level of pollution, e.g. the rising level of CO_2 in the atmosphere

[2] We leave out all issues related to opinion formation and political debate within a country but focus on the negotiations taking place at the global level.

[3] For example, Battaglini and Harstad (2012) interpret x_i to be the level of energy used to produce some good. For simplicity we assume one unit of consumption to generate one unit of pollution.

that significantly contributes to global warming. In addition to the global impact, effects of emissions differ locally and are experienced by a certain subgroup of countries. For instance, short-lived climate pollutants such as black carbon, methane and tropospheric ozone have both a local and global impact.[4]

We model the local spillover effect by a network structure $g \in G^N$, where $G^N = \{g \mid g \subseteq g^N\}$ denotes the set of all possible networks on the set of players N, with g^N denoting the set of all subsets of N of size 2. A link between two countries in the network then describes the presence of a direct local spillover which could be due to geographical distance, common borders, sharing an ocean or a lake, or other underlying assumptions that we exclude from our model. We assume that local emission spillovers between countries are bidirectional and thus focus on undirected networks, however we show in Section 7 that this is not restrictive and adapting notation our results also hold for directed networks. Furthermore, we leave out scaling issues of the effects and consider only unweighted graphs.

The network structure is captured via the indicator function g_{ij} which is equal to 1, if i and j are neighbors and 0 in all other cases. With respect to the spillover effect, every country suffers from its own emissions both through the local and the global effect. To account for this and to incorporate it into our model, we let $\bar{g}_{ij} = g_{ij}$ for all $i \neq j$ and $\bar{g}_{ii} = 1$.

We assume linear spillover effects on both the global and local level due to analytical tractability. The marginal impacts are weighted relative to benefits from consumption by $\beta > 0$ for the global spillover effect and $\gamma > 0$ for the local spillover effect.[5] The costs incurred from total pollution are then represented by the cost function

$$K_i(x_i, x_{-i}) = \beta \sum_{j \in N} x_j + \gamma \sum_{j \in N} \bar{g}_{ij} x_j.$$

The individual profit π_i of a country $i \in N$ can thus be represented as follows:

$$\pi_i\left(x_i, x_{-i}\right) = B_i(x_i) - K_i(x_i, x_{-i}) = -\frac{1}{2}\left(\bar{x}_i - x_i\right)^2 - \beta \sum_{j \in N} x_j - \gamma \sum_{j \in N} \bar{g}_{ij} x_j. \qquad (1)$$

In line with the standard literature, we will represent an IEA that is formed among countries by the game-theoretic concept of a coalition. More specifically, we denote by $C \subseteq N$ the coalition of k countries $i_1, \ldots, i_k$ that cooperate on the abatement level to maximize the utilitarian welfare of its members.[6] A member of a coalition, called

[4] Also, usually not only one single pollutant is released during production or consumption but others are emitted simultaneously and these might only impact certain, local areas. We abstract from this by summarizing all different pollutants in one representative emission flow x_i.

[5] We shall mention that we abstract from heterogeneities with respect to marginal impacts to focus on the effect that is derived from the network position.

[6] In this benchmark model we abstract from the possibility of multiple agreements, thus only one coalition can form even though the consideration of a local spillover structure may naturally induce locally organized agreements and thus multiple coalitions. We will come back to this in our discussion in Section 8.

signatory, hence chooses a pollution level x_i such that it maximizes the sum of all signatories' utility. Given a coalition C, we denote by $C+i$ the coalition when i joins C. A main driver of our results will be the number of intra-coalition links, i.e. the number of neighbors that are part of the coalition, which will be denoted by $k_i = |N_i \cap C|$.

3.2 The Free-Rider Problem in the Single Stage Game

To illustrate the issues that arise when forming IEAs, we first look at the two extreme cases of either no or full cooperation, i.e. $C = \emptyset$ and $C = N$. In the one-shot game all countries simultaneously choose their level of emissions x_i.[7]

In the situation of no cooperation, i.e. $C = \emptyset$, every country myopically determines its emission level to maximize individual profit π_i as defined in (1). The first order conditions (subsequently abbreviated as FOCs) then directly yield the *non-signatory* Nash outcome

$$x_i^{NS} = \bar{x}_i - \beta - \gamma. \tag{2}$$

Hence, in absence of an IEA, every country emits just slightly below its first-best level $\bar{x}_i$ by accounting for the own marginal emission effect $\beta + \gamma$.

Assumption 1. As we assume non-negative emissions, we impose the following condition for all $i \in N$: $\bar{x}_i \geq m_1\beta + m_2\gamma,\ \forall\ 0 \leq m_1, m_2 \leq n$.

We think of an IEA such that members choose emissions to maximize the utilitarian welfare restricted to the members of the IEA. Denoting by $C \subseteq N$ the set of countries who sign the IEA, the maximization problems for those countries, hence, are given by,

$$\max_{(x_i)_{i \in C}} \sum_{i \in C} \pi_i \left(x_C, x_{N \setminus C} \right). \tag{3}$$

The FOCs yield an optimal emission level for every signatory that depends on the size of the coalition, k, and the number of intra-coalition links, $k_i = |N_i \cap C|$,

$$x_i^S(C) = \bar{x}_i - \beta k - \gamma(k_i + 1), \tag{4}$$

such that all non-signatories $j \notin C$ choose the emission level x_j^{NS} given by (2). In the following, we will denote by $x(C) = \left((x_i^S(C))_{i \in C}, (x_j^{NS})_{j \in N \setminus C} \right)$ the vector of outputs when a coalition C is collaborating and denote profits by $\pi_i(C) = \pi_i(x(C))$.

For the case of full, global cooperation, i.e. $C = N$, utilitarian welfare of all countries $\sum_{i \in N} \pi_i(x)$ is maximized. We get, $x_i^S(N) = \bar{x}_i - \beta n - \gamma\left(\eta_i + 1\right)$, where every country takes into account the global effects from its pollution as well as the local spillovers to every respective neighbor. From a global perspective, this would be the first-best solution

[7] The assumption of simultaneous move is standard in the literature. There are, however, also papers that study the effects of a coalition that acts as a Stackelberg leader (e.g., Rubio and Ulph, 2006).

as all externalities are internalized. However, not every individual country is necessarily better off under full cooperation than under no cooperation. To demonstrate this, let us denote by $\Delta_i \left(N, \emptyset\right)$ the difference in individual profits between global an no cooperation. We have that

$$\Delta_i \left(N, \emptyset\right) = \pi_i(N) - \pi_i(\emptyset) = \frac{1}{2}\beta^2 \left(n-1\right)^2 + \gamma \sum_{j\neq i}(\gamma g_{ij} + \beta)\eta_j - \frac{1}{2}\gamma^2\eta_i^2.$$

Thus, the potential gains from a full cooperation agreement are positive if and only if

$$\gamma^2(\eta_i^2 - \sum_{j\in N} g_{ij}\eta_j) \leq \beta^2(n-1)^2 + 2\gamma\beta\sum_{j\neq i}\eta_j. \qquad (5)$$

This condition obviously holds if either the network is not too asymmetric (i.e. where $\eta_i \simeq \eta_j \ \forall i, j \in N$) or in those cases where the global impact β is large compared to the local impact γ.[8] Instead, very asymmetric network structures and a high local spillover effect γ can lead to cases where countries actually prefer no cooperation to full cooperation as we see in Example 1. Thus the first observation that we can take away here is that the local spillover structure may yield asymmetries between countries incentives which are difficult to overcome when forming IEAs.

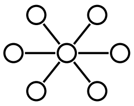


Fig. 3.1: A star network with one central node and 6 peripheral nodes.

Example 1. Consider the star network $g^*(n)$ with one player connected to all other $n-1$ players who are only connected to the center, exemplarily shown in Figure 3.1 for 7 nodes. The center node prefers no cooperation over full cooperation for all values of $\gamma \geq \beta\frac{n-1}{n-3}$. For the peripheral nodes, no cooperation is clearly worse than the full-cooperative solution. Nevertheless, global cooperation is not a Pareto-improving outcome to no cooperation.

So far, we have only compared global cooperation, when every country chooses emission as to maximize utilitarian welfare, to no cooperation. As the members jointly maximize their profits, each signatory takes into account the coalition's global spillover effect as well as all the local spillover effects from signatories in the neighborhood. Thus, the signatories' strategies are efficient within the coalition C with respect to the utilitarian welfare function. However, because of the lack of a central agency, an IEA, if formed, has to be *self-enforcing* in order to become effective. That is, no member shall have an incentive to join or to leave an IEA. In other words, formation of an IEA follows an

[8] Two examples of sufficient conditions for (5) to hold are that either the network is regular, i.e. $\eta_i = \eta_j$ for all $i, j \in N$ or that $\gamma \leq \beta$.

open membership single coalition game (for more details see e.g., Finus and Rundshagen, 2001; Yi and Shin, 2000). Countries voluntarily decide whether to become member of the IEA or not and no country can be excluded from membership nor can any member be forced to stay in the coalition. Given an IEA C, the incentive for a signatory $i \in C$ to leave can be calculated to be,[9]

$$\begin{aligned}\pi_i(C) - \pi_i(C \setminus \{i\}) &= -\frac{\beta^2}{2}k^2 - \frac{\gamma^2}{2}k_i^2 - \beta\gamma k k_i + 2\beta^2 k + \gamma(3\beta+\gamma)k_i - \frac{3}{2}\beta^2 \\ &= -k_i\gamma\left(\beta(k-3) + \gamma(\tfrac{k_i}{2} - 1)\right) - \tfrac{\beta^2}{2}(k-3)(k-1). \qquad (6)\end{aligned}$$

Thus, the structure of local spillovers plays a crucial role in each country's decision to whether or not join a coalition. In particular, no IEA C of $k > 3$ signatories can contain a member i with $k_i \neq 1$ neighbors in C since otherwise (6) becomes negative and hence there is an incentive to leave this IEA. For the IEA including all countries this means that even those countries that satisfy condition (5) would rather free-ride on the others' efforts than choosing the full-cooperative output level, if the network is large enough.

Proposition 1. *Let $n > 3$. For all countries $i \in N$ with $\eta_i \neq 1$ it holds that*

$$\pi_i(N \setminus \{i\}) > \pi_i(N).$$

For those countries with only one neighbor, i.e. $\eta_i = 1$, the above holds if and only if

$$\gamma < \beta\left(n - 3 + \sqrt{(n-3)(2n-4)}\right).$$

Hence if and only if the network consists of separated pairs only and the local spillover effect is significantly larger than the global spillover effect, global cooperation may be a Nash equilibrium which implies that the IEA where all countries sign is stable. In general however we can conclude that in the networks we consider the full cooperative outcome is never a Nash equilibrium of the single-stage game and no country would choose the full-cooperative output level unilaterally but rather free-ride on the others' efforts.

4 Stable IEAs in Infinitely Repeated Games

As the nature of pollution and production is rather of repeated form, a one-shot game may not be the accurate model to consider IEAs. So while in the one-shot game IEAs, particularly those consisting of many countries, fail to be self-enforcing or stable, there is hope for stability of IEAs by threat of future punishment in the repeated game. The main question that we ask here is which IEAs are implementable when the nature of game is of repeated form. We do not ask how these IEAs form, but rather whether certain IEAs are stable and which conditions on the local spillover structure foster or

[9] See a supplementary appendix online, available at https://sites.google.com/site/guenthermichael/.

harm implementability of an IEA. We informally discuss the formation of an IEA in our extensions in Section 7.

Similarly to the one-shot case, we model an IEA in the infinitely repeated game by a coalition of signatories with two stability conditions. First, stability requires that no signatory has an incentive to deviate from the strategy that maximizes the coalition's utilitarian welfare through threats of future punishments by the other signatories. These threats deter free-rider incentives and allow for the implementation of full cooperation as a subgame perfect equilibrium (subsequently abbreviated as SGP equilibrium) as long as the discount factor is high enough (see Fudenberg and Maskin, 1986). Second, stability requires execution of the punishment strategies such that they are not vulnerable to renegotiation. In other words, the punishers shall not have an incentive to *renegotiate* the terms of the agreement and restart the game.

We rule out this possibility by considering only those equilibria as stable outcomes that are renegotiation-proof. More specifically, we apply the most frequently used notion of weak renegotiation-proofness (subsequently abbreviated as WRP).[10]

4.1 The Infinitely Repeated Game

We briefly introduce a standard infinitely repeated game of the stage game as described in Section 3. Time is discrete and indexed by $t \in \mathbb{N}$. In each period, countries choose consumption (i.e. emission levels) $x_i(t)$ (with slight abuse of notation). In other words, the stage game is played in each period. At time t, country i's choice of emission may depend on the entire history of the game through period $t-1$, denoted $h^{t-1} = \Big((x_1(1), ..., x_n(1)), ..., (x_1(t-1), ..., x_n(t-1))\Big)$. Thus, a strategy $\mathbf{s_i}$ for country i is a function that, for every date t and every possible history h^{t-1}, defines a period t action $x_i(t) \in \mathbb{R}_+$. Future payoffs are discounted with a common discount factor $\delta < 1$ such that each country i receives a discounted payoff for a sequence of emissions $\big\{(x_i(t), x_{-i}(t))\big\}_{t=0}^{\infty}$,

$$\Pi_i = (1-\delta)\sum_{t=0}^{\infty} \delta^t \pi_i(x_i(t), x_{-i}(t)). \tag{7}$$

In the infinitely repeated game, a weakly renegotiation-proof equilibrium is defined as a strategy profile of the repeated game $\mathbf{s} = (\mathbf{s_i})_{i \in N}$ such that it satisfies the following.

[10] Note that there are two limitations of this concept in our subsequent analysis: First, weak renegotiation-proofness takes account of the possibility of a unilateral deviation of a single country but does not regard the possibility of a deviation of a subset of countries, which may very well occur as a result of coordinated action among some countries. Second, by deriving a WRP equilibrium we can not answer the question of how coordination may be achieved, i.e. how countries agree on a particular IEA (see also the discussion in Asheim and Holtsmark, 2009).

Definition 1. [Farrell and Maskin (1989)] A simple strategy profile $\mathbf{s}$ is a *weakly renegotiation-proof (WRP) equilibrium* of the infinitely repeated game if and only if (1) $\mathbf{s}$ is a subgame perfect equilibrium of the infinitely repeated game and (2) there exist no two continuation equilibria such that all players strictly prefer the one to the other.

As we view an IEA as a coalition of players $C \subseteq N$ that should implement the signatory strategy x_C^S as defined by (4), the application of the concept of WRP equilibrium has exactly the conditions desired for our setup.[11] Part (1) of Definition 1 ensures that the coalition is stable with respect to deviations and (2) implies stability with respect to renegotiations. Hence, we ask whether the signatory emission can be supported by a WRP equilibrium of the repeated game. To be implementable, *credible* punishment paths have to be designed to deter deviations. Due to renegotiation incentives, it may not be optimal that all other signatories punish. In fact, the harsher the punishment or the more countries punish, the higher is the incentive to renegotiate.[12] Another aspect of implementability of an IEA is the fact that strategies shall be simple. We use the notion of simple strategies by Abreu (1988). That is, we focus on punishments that last only one period, the set of punishers is time-invariant and all punishers use the same punishment action.

We denote $P_i(g) \subseteq C$ the set of players that punish deviator $i \in C$. Punishment for a deviating signatory $i \in C$ is thus carried out as follows: each country $j \in P_i$ punishes a deviation of country i by emitting the punishment level x_i^P instead of the signatory emission x_i^S where

$$x_j^P = \bar{x}_j - p(\beta + \gamma)$$

in the period after the deviation. We assume that $p \geq 1$ such that the highest punishment level is the Nash output x_j^{NS} (in Subsection 7.1 we relax this assumption). As all non-signatories $l \notin C$ play their first-best action, i.e. the Nash equilibrium level x_l^{NS}, they will not punish a deviator but continue with their strategy.

A coalition $C = \{i_1, \dots, i_k\}$ is then implemented through a strategy $\mathbf{s^C}$, which is defined for period $t = 1$ by $\mathbf{s_i^C}(\emptyset, 1) = x_i^S$ for all $i \in C$ and for $t \in \mathbb{N}$ recursively defined by

$$\mathbf{s_i^C}(h^{t-1}, t) = \begin{cases} x_i^P, & \text{if } \exists! j \in C : x_j(t-1) \neq \mathbf{s_j^C}(h^{t-2}, t-1) \text{ and } i \in P_j \\ x_i^S, & \text{else} \end{cases}.$$

[11] Note that in the following we will use the notation $x_i^S \equiv x_i^S(C)$ unless otherwise stated. Also, to shorten notation we will omit the vector $x_{N\setminus C}$.

[12] It has been shown in other papers, such as Froyn and Hovi (2008) and Asheim et al. (2006), that the limitation of punishers to a subset of the cooperating players can decrease the incentives for renegotiation. The same effect can be observed in our model. In our model though, players face heterogeneous costs from pollution through the local spillover channel. Thus, we have to derive individual punishment paths and therefore individual sets of punishers for any possible deviator. Whereas for example Asheim et al. (2006) artificially introduce two separated regions and let a deviating country be punished only by countries in the same region as the deviator, we allow for more flexible punishment sets and focus on the impact of the local spillover structure, represented by the network g, on possible equilibrium outcomes.

Since non-signatories $j \in N \setminus C$ stick to their Nash emission level x_j^{NS} throughout the game, their strategy is simply given by $\mathbf{s_j^{N\setminus C}}(\cdot, t) = x_j^{NS}$ for all $t \in \mathbb{N}$. This defines a simple strategy profile in the spirit of Abreu (1988), since it gives rise to the $(n+1)$–vector of paths

$$(\mathbf{a^C}, \mathbf{p_1^C}, \dots, \mathbf{p_n^C}).$$

where $\mathbf{a^C}$ is the agreement path, s.t.

$$\mathbf{a^C} = \left\{ \left(x_C^S, x_{N\setminus C}^{NS} \right), \left(x_C^S, x_{N\setminus C}^{NS} \right), \dots \right\}$$

and the punishment paths which are triggered if a single country $i \in N$ deviates,

$$\mathbf{p_i^C} = \left\{ \left(x_{P_i}^P, x_{C\setminus P_i}^S, x_{N\setminus C}^{NS} \right), \left(x_C^S, x_{N\setminus C}^{NS} \right), \left(x_C^S, x_{N\setminus C}^{NS} \right), \dots \right\}$$

In other words, any single deviation of a country i results in a one-period punishment by the countries P_i who subsequently revert to their signatory strategies, while all others play as in the agreement path.[13] Moreover, only signatories may have an incentive to deviate and thus need to be punished – non-signatories will always stick to their Nash output and thus will not be punished directly. All together we ask whether $\mathbf{s} = (\mathbf{s^C}, \mathbf{s^{N\setminus C}})$ forms a WRP.

It is worth remarking that we consider a very specific punishment rule and simple punishment strategies. However, as stated before, it is a strategy that is simple to implement and therefore suitable for the application in an IEA. Moreover, it is the one that has been frequently used in the repeated games literature on IEAs and our results show that this may not be sufficient to establish full cooperation as a WRP equilibrium. In Subsection 7.1 we also briefly discuss what changes if we allowed for other punishment strategies.

4.2 Weakly Renegotiation-Proof Coalitions

Since implementability of an IEA C depends on the existence of an WRP equilibrium supporting the C optimal punishments sets P_i and punishment level p have to be determined for each $i \in C$. As spillovers are heterogeneously distributed across the signatories due to their network position, this might be a quite complex task. Necessary and sufficient conditions on these punishment sets are presented in the following result.

[13] Note that only single deviations are considered, that is multiple deviations in a single period are not punished.

Theorem 1. *The simple strategy profile* $\mathbf{s}$ *that implements the coalition* C *is a WRP equilibrium of the repeated game if and only if for all* $i \in C$

$$\delta \left[\beta^2 |P_i|(k-p) + \beta\gamma \left(|P_i|(1-p) + \sum_{m \in P_i} k_m + |P_i \cap N_i|(k-p) \right) \right.$$
$$\left. + \gamma^2 \left(\sum_{m \in P_i \cap N_i} k_m + |P_i \cap N_i|(1-p) \right) \right] - \frac{1}{2} \big(\beta(k-1) + \gamma k_i\big)^2 \geq 0 \quad (8)$$

and for all $i \in C$ *there exists at least one* $j \in P_i$ *such that*

$$\beta^2(k-p)(|P_i|-p) + \beta\gamma \left((|P_i|-p)(1-p) + \sum_{m \in P_i \setminus \{j\}} k_m + |P_i \cap N_j|(k-p) \right)$$
$$+ \frac{\gamma^2}{2} \left(2 \sum_{m \in P_i \cap N_j} k_m + \Big(2|P_i \cap N_j| + 1 - p\Big)(1-p) \right) - \frac{1}{2} \big(\beta(k-1) + \gamma k_j\big)^2 \leq 0. \quad (9)$$

First, Equation (8) yields the condition for $\mathbf{s}$ to be a subgame perfect equilibrium. In particular P_i needs to be large enough while p must be low enough in order for Equation (8) to hold.[14]. While punishment needs to be harsh enough to deter deviations, Equation (9) specifies conditions for weak renegotiation-proofness. In particular, punishment cannot be too harsh to prevent incentives for renegotiation.[15]

Hence, if both conditions are satisfied, the simple strategy profile that specifies for every signatory $i \in C$ a set of punishers P_i and a punishment level of the punishers p sustains the coalition C as a WRP equilibrium. Even with the simple strategies that we consider here, the conditions in Theorem 1 seem quite complex. The complexity stems from the heterogeneous spillover channels represented by the network. The intuition of the conditions can be best explained in the following when we focus only on one type of spillover (global respectively local, Section 4.3) and subsequently explore comparative statics with respect to changes in the network and punishing sets (Section 4.4).

In order to better understand the conditions of Theorem 1 with respect to the specific punishment sets and the network, in the following we consider special cases of spillover

[14] Note that the SPE condition would also entail that no player $j \in P_i$ has an incentive to unilaterally not carry out his punishment. This however is automatically satisfied as we assume $p \geq 1$ (see also Lemma 2 in the Appendix)

[15] Note that Equation (9) only has to hold for one element of the punishment set which may lead to results such that enlargement of the punishment group may actually benefit conididiton (8). However, the results presented in this paper also hold for stronger versions of WRP (e.g. that (8) has to hold for all $j \in P_i$), which are not available in the literature so far.

and network structures. For the rest of the paper we will moreover assume that every country punishes a deviation by emitting its Nash output level, i.e. for all $i \in N$ we set $p = 1$ for all punishers $j \in P_i$.

4.3 WRP Conditions for Special Spillover Structures

First, suppose that there exists only the global spillover channel, as e.g. in Asheim and Holtsmark (2009). Hence, the underlying network plays no role and the only heterogeneity in the game stems from the exogenously given satiation levels $\bar{x}_i$. However, as these do not influence the results, the intuition alone implies that it is not important who punishes, but how many punish, i.e. it is not the composition of the punishment set that matters but the size. Indeed, setting $\gamma = 0$ in (8) and (9), one obtains the following.

Corollary 1. *For $\gamma = 0$, the conditions of Theorem 1 reduce to*

$$\frac{1}{2\delta}(k-1) \leq |P_i| \leq \frac{1}{2}(k+1) \qquad \forall i \in C.$$

Without the local spillover effect, the conditions of Theorem 1 determine the number of punishers allocated to each signatory in order to be part of a WRP coalition. To give some intuition, the punishment set needs to be large enough in order to deter deviation (first inequality) while it cannot be too large in order to prevent renegotiation (second inequality). The conditions of Corollary 1 are equivalent to the conditions of Asheim and Holtsmark (2009), Theorem 1, with $s = k$ and $p = 1$.

Second, if we instead consider general spillover effects but very special networks, then similar observations can be made. For example the empty network $g = g^{\emptyset}$ is trivially equivalent to the case where no local spillover effects exist. Further, consider $g = g^N$, i.e. the complete network. Then, all countries experience the local spillover from a given country which immediately implies that this is equivalent to the case where the magnitude of the global spillover is $\beta + \gamma$ while there are no local spillovers. Hence, also for the case of the complete networks, the conditions for a WRP equilibrium are equivalent to the ones from Corollary 1.

Third, consider the case when the global spillover channel does not exist. Then, the game boils down to a local spillover game where countries can only free-ride on the actions of their direct neighbors.

Corollary 2. *Let $\beta = 0$. The conditions of Theorem 1 on the punishment set P_i, $i \in C$ reduce to the following:*

$$\begin{aligned} &\forall\ i \in C & \sum_{m \in P_i \cap N_i} k_m \geq \tfrac{k_i^2}{2\delta}, \\ &\forall\ i \in C,\ \exists\ j \in P_i,\ s.t. & \sum_{m \in P_i \cap N_j} k_m \leq \tfrac{k_j^2}{2}. \end{aligned}$$

When only the local externalities play a role, then the composition of the punishment sets becomes important. Since only the local spillover channel is present, only neighbors have a deterring effect. The first condition requires punishment to be harsh enough in order to deter deviation, which implies that the set of punishers must have enough neighbors in C. This is due to the fact that emission is increased from $x_j^S = \bar{x}_j - k_j\gamma$ to $x_j^P = \bar{x}_j - \gamma$ if $j \in P_i$. With the same reasoning, the incentive to deviate for country i is determined by k_i. On the other hand, total punishment must not be too harsh in the sense that the punishers shall not have an incentive to renegotiate, which is presented in the second condition. Incentives to renegotiate occur if the neighborhood structure of the punishers overlaps too much. Thus, given a punishment level, the set of punishers should be constructed such that their local spillover channels interfere minimally.

4.4 Comparative Statics

Abstracting from the special cases of only global respectively local spillovers, we further explore the meaning of the conditions for existence of a WRP equilibrium in simple strategies to support a coalition $C \subseteq N$ given in Theorem 1 by means of comparitive statics. To understand the effect of the different spillover channels, i.e. the underlying network, we study the effect of additional links in the network on the conditions of subgame perfection and weakly renegotiation-proofness. Further, we ask how an enlargement of the punishment group may impact these conditions, or more precisely, what the marginal effect of an additional punishing country is.

4.4.1 The effect of the spillover structure

First, consider the condition on subgame perfection (see Theorem 1, Equation 8). Take $i \in C$ and define the function $f_i(\delta, C, g, P_i)$ as the left-hand side of (8). The marginal effect of an additional link (currently not in the network) $lm \notin g$, $l, m \neq i$, on the subgame perfect condition of player i can be calculated to be,

$$f_i(\delta, C, g + lm, P_i) - f_i(\delta, C, g, P_i) = \left(\mathbb{1}_{P_i}(l) + \mathbb{1}_{P_i}(m)\right) \delta \Big(\beta\gamma \left(\mathbb{1}_{N_i}(l) + \mathbb{1}_{N_i}(m)\right) \Big),$$

where $\mathbb{1}_A(i)$ denotes the indicator function such that $\mathbb{1}_A(i) = 1$ if $i \in A$ and $\mathbb{1}_A(i) = 0$ else. The marginal effect is positive as long as the link lm involves at least one of i's punishers ($\mathbb{1}_P(l) = 1$), meaning that condition (8) is more likely to hold for $i \in C$ after link addition since (8) requires $f_i(\delta, C, g, P_i) \geq 0$. Thus, the marginal effect of an additional spillover channel is largest if the link is between two punishers of i who are also neighbors with i and lowest if both are neither. The reason is that punishment increases if a punisher has an additional spillover channel, since emission reduction is higher in the non-punishment case. Moreover, neighbors cause a larger marginal effect for country i through the additional spillover channel.

While the marginal effect of link addition between two countries other than i on i's incentive to play the signatory strategy is unambiguously non-negative, the same is not so clear for the marginal effect if i itself is involved in the additional link:

$$f_i(\delta, C, g+im, P_i) - f_i(\delta, C, g, P_i) = \beta\gamma\Big(\mathbb{1}_{P_i}(m)\delta k - (k-1)\Big) + \frac{\gamma^2}{2}(\mathbb{1}_{P_i}(m)2\delta - (2k_i+1)).$$

Obviously, if a deviator i has additional spillover channels to non-punishers, i.e. $\mathbb{1}_{P_i}(m) = 0$, the effect is negative, since i is required to reduce more of its emission in the signatory strategy, and, thus, more tempted to deviate. If instead the additional link is to a punishing player, i.e. $\mathbb{1}_{P_i}(m) = 1$, then punishment also increases. This has an additional deterring effect which clearly depends on δ such that the effect on subgame perfection is negative as long as the discount factor δ is small enough, i.e. $\delta \leq \bar{\delta}(g) = \frac{k-1+\frac{\gamma}{\beta}(k_i+\frac{1}{2})}{k+\frac{\gamma}{\beta}}$. Note that the marginal effect is negative for all discount factors if $\bar{\delta}(g) \geq 1$, which holds for large enough k_i and marginal local spillovers γ.

Next, we turn to the second condition of stability of Theorem 1, i.e. the condition that ensures weak renegotiation-proofness. Considering a deviator $i \in C$ and a punisher $j \in P_i$, we define $h_{ij}(\delta, C, g, P_i)$ as the left-hand side of (9). The marginal effect of an additional link $lm \notin g$ on the incentives of a punisher j of deviator i is then given by

$$h_{ij}(\delta, C, g+lm, P_i) - h_{ij}(\delta, C, g, P_i) = \big(\mathbb{1}_{P_i}(l) + \mathbb{1}_{P_i}(m)\big)\Big(\beta\gamma\left(\mathbb{1}_{N_i}(l) + \mathbb{1}_{N_i}(m)\right)\Big).$$

Since the marginal effect is positive if at least one link involves a punisher of i, condition (9) is less likely to hold for $j \in C$ after link addition since (9) requires $h_{ij}(\delta, C, g, P_i) \leq 0$. The marginal effect of an additional link between l and m on the incentives of $j \in P_i$ to renegotiate is largest, when both l and m are punishers and neighbors of j.[16] If there is an additional link between two countries that are not in the punishing group P_i, this has obviously no impact on the WRP condition for j. Thus, an overlapping spillover structure of punishing group P_i makes renegotiation more attractive (and thus makes the coalition vulnerable to renegotiation) since the profit under cooperation increases. This effect is fostered if there is also a connection to the punisher j, as decreasing costs through local spillovers increase j's incentives to renegotiate and not carry out the punishment.

Since we study the incentives for j to renegotiate, it makes a difference if j itself is part of the additional spillover. We obtain

$$h_{ij}(\delta, C, g+jm, P_i) - h_{ij}(\delta, C, g, P_i) = \begin{cases} -\beta\gamma(k-1) - \gamma^2(k_j + \frac{1}{2}), & m \notin P_i \\ \beta\gamma - \gamma^2(k_j - k_m - \frac{1}{2}), & m \in P_i \end{cases}.$$

For j's incentive itself to renegotiate, the effect of additional links is ambiguous. First, if the additional link leads to a non-punisher of i, then j has lower benefits from cooperation compared to her Nash strategy making renegotiation less attractive. If instead

[16] Note that here $m = i$ is not excluded. But since i cannot be part of the punishment group, we always have $\mathbb{1}_{P_i}(i) = 0$.

the additional link is to a punisher of i, then j also suffers from the punishment level of the additional neighbor during the punishment phase, which harms j through the local spillover channel and hence works in the opposite direction to make renegotiation more attractive.

We can conclude: Given the punishment group, higher density of the spillover structure within the coalition facilitates subgame perfection while it harms the renegotiation-proofness condition for most countries. Note, however, that this does not necessarily hold for the potential deviator respectively a potential punisher who is involved in the additional link. So while the subgame perfection condition (8) has to hold for all $i \in C$ and renegotiation-proofness condition (9) for at least one $j \in P_i$, the overall effect of link addition on both stability conditions of an IEA may be ambiguous.

4.4.2 Additional punishers

In order to determine an individual, optimal punishment group for every possible deviator of a coalition, we also have to understand what is the marginal effect of an additional punisher for the two conditions for WRP coalitions. First, we study the effect on the SGP condition (8). We have the following marginal effect of an additional punisher l on the deviator i:

$$f_i(\delta, C, g, P_i \cup \{l\}) - f_i(\delta, C, g, P_i) = (\beta + \mathbb{1}_{N_i}(l)\gamma)(\beta(k-1) + \gamma k_l).$$

Obviously, any additional punisher will increase the deterring effect on a deviator and if the punisher is a neighbor, then the additional spillover channel gives rise to a larger effect. For the WRP condition (9), we have the following marginal effect of an additional punisher l on a punisher j:

$$h_{ij}(\delta, C, g, P_i \cup \{l\}) - h_{ij}(\delta, C, g, P_i) = (\beta + \mathbb{1}_{N_j}(l)\gamma)(\beta(k-1) + \gamma k_l).$$

Here, the more punishers the higher incentives to renegotiate, again the effect is enhanced if the punishers are also neighbors.

The comparative statics have shown that there is a trade-off in characterizing the optimal punishment group for each coalition-member: the more punishers and the higher the connectedness among them, the higher the threat of punishment and the easier to sustain an SGP equilibrium. In turn, incentives to renegotiate increase with the size of the punishment group and its clustering. In the next section we will study how to find a punishment group for each signatory to sustain a coalition as a WRP equilibrium.

5 The Stability of a Global IEA

Having determined general conditions for the stability of an IEA, the question of existence of such a sustainable IEA has not yet been answered. We focus here on the stability of a worldwide IEA, i.e. an IEA where every country plays the signatory strategy and has no incentive to deviate or renegotiate. However, it is rather obvious that not all network structures allow for a WRP equilibrium that supports a global IEA. For instance, from Example 1 we already know that the center player of a star network prefers no cooperation to global cooperation if the local externality is large enough, i.e. $\gamma \geq \beta \frac{n-1}{n-3}$. This immediately implies that a subgame perfect equilibrium supporting global cooperation cannot exist in the repeated game for $\gamma \geq \beta \frac{n-1}{n-3}$. Clearly, adding the WRP condition (9) to this, makes the existence of a stable global cooperation even more restrictive. In fact it can be shown that for large star networks, a WRP supporting global cooperation fails to exist.

Proposition 2. *Consider the star network $g^*(n)$ and let γ and β be independent of n. Then for $n \to \infty$, there does not exist a WRP equilibrium in simple strategies supporting the global IEA.*

The intuition behind this result is that as n grows, the players become more and more heterogeneous in terms of degree. While the center player of the star has to reduce his emission increasingly in the number of his neighbors, the set of punishers has to grow as well in order to deter deviation by the center player. This, however, gives increasing incentives to renegotiate – implying non-existence of a WRP.[17] This is a fundamental difference to the existing models in the literature that study the possibility of global cooperation as a stable outcome of the climate game. Unlike in Asheim and Holtsmark (2009), we have shown that for very asymmetric networks such as the star network, global cooperation may fail to be a WRP coalition for the specific punishment rule.

Hence, it seems to be the asymmetry of the network – in particular the asymmetry of degrees – which leads to failure of a global IEA. Instead, we may also look at the other extreme case of spillover networks where there are not heterogeneities in terms of degree, i.e. a network where the number of neighbors of all players are the same. Such a structure is defined as a *regular* network.

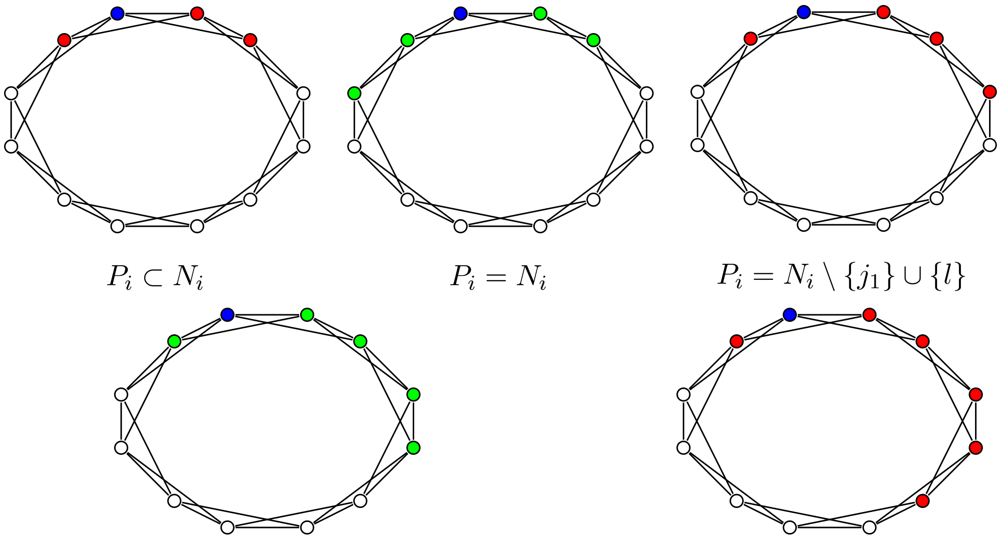
Example 2. Consider a regular network of $n = 12$ players with $\eta_i = 4 \; \forall \; i \in N$, illustrated in Figure 5.1. Let $\beta = \gamma$ and for simplicity $\delta \to 1$. Which punishment sets P_i sustain full cooperation as a WRP coalition?

The conditions for the grand coalition to be a WRP equilibrium (cf. Theorem 1) then read:

$$|P_i| + |P_i \cap N_i| \geq 7.5 \; \forall \; i \in N, \tag{10}$$

$$|P_i| + |P_i \cap N_j| \leq 8.5 \; \forall \; j \in P_i, \; \forall \; i \in N. \tag{11}$$

[17] This result is not restricted to the star network only. Suppose the parameter setting is such that we can sustain full cooperation as a WRP coalition in the star network. Then, as seen in the comparative statics section above, the addition of one single link may change the marginal incentives such that the punishment structure needs to be redesigned and may end up to not sustain full cooperation as a WRP coalition. For instance, in the 5-player star network with $\gamma = 2\beta$, full cooperation is a WRP equilibrium. However, if two peripheral nodes are linked, this is no longer the case.



$P_i \subset N_i$ $P_i = N_i$ $P_i = N_i \setminus \{j_1\} \cup \{l\}$

$P_i = N_i \setminus \{j_1\} \cup \{l_1, l_2\}$ $P_i = N_i \setminus \{j_1\} \cup \{l_1, l_2, l_3\}$

Fig. 5.1: Different scenarios in a circle network.
Green nodes represent a WRP punishing set.

Denote the neighbors of player i by $N_i = \{j_1, j_2, j_3, j_4\}$. While for $P_i \subset N_i$, $P_i \neq N_i$, the punishment can be calculated to be too low, choosing $P_i = N_i$ both conditions (10) and (11) are satisfied:

$$\begin{aligned}(10) &\Leftrightarrow 4 + 4 = 8 > 7.5,\\ (11) &\Leftrightarrow 2 + 3 = 5 < 8.5.\end{aligned}$$

Another possible punishment set giving rise to a WRP equilibrium supporting the global IEA can be calculated to be $P_i = N_i \setminus \{j_1\} \cup \{l_1, l_2\}$ with $l_1, l_2 \notin N_i$. However, punishment sets like $P_i = N_i \setminus \{j_1\} \cup \{l_1\}$ with $l_1 \notin N_i$ do not satisfy (10) while $P_i = N_i \setminus \{j_1\} \cup \{l_1, l_2, l_3\}$ with $l_1, l_2, l_3 \notin N_i$ do not satisfy (11). The different scenarios are displayed in Figure 5.1.

While for very asymmetric network structures WRP equilibria supporting the global IEA fail to exist (cf. Proposition 2), a symmetric network structure as given in Example 2 allows for stability of the global IEA. If symmetry in the degree is given, these findings can indeed be generalized by considering regular networks such that $\eta_i = \eta_j = \eta$ for all $i, j \in N$.

Focusing on the global IEA in regular networks yields that $k_i = \eta$ for all $i \in N$, implying that conditions (8) and (9) then simplify such that

$$\delta \left(|P_i \cap N_i| \gamma + |P_i| \beta \right) \geq \tfrac{1}{2} \left(\beta(n-1) + \gamma\eta \right) \ \ \forall\, i \in N \tag{12}$$

$$\left(|P_i \cap N_j| \gamma + (|P_i| - 1)\beta \right) \leq \tfrac{1}{2} \left(\beta(n-1) + \gamma\eta \right) \ \ \forall\, j \in P_i, \ \forall\, i \in N. \tag{13}$$

Using the results from the comparative statics analysis, we can now directly derive the main result of this section.

Proposition 3. *Let $\delta \to 1$. Then, for every regular network there exists a WRP equilibrium supporting the global IEA.*

This result underlines the intuition that in very symmetric settings it is easier to achieve cooperation than in very asymmetric network settings such as the star. More specifically, this proposition yields that if countries are sufficiently patient, we can always find a punishment set P_i such that the simple strategy profile $\mathbf{s}$ sustains the grand coalition as a WRP equilibrium of the infinitely repeated game.[18]

Note that the condition $\delta \to 1$ is not a binding condition. It is impossible to consider all possible punishment sets for all possible regular networks, but it is easy to argue that the complete network is the most restrictive case, since there clustering is 1 and all spillover channels are present. For the complete network, the threshold value for the discount factor δ can be easily determined and coincides with the the threshold in Asheim and Holtsmark (2009) since full cooperation can be established as a WRP equilibrium if the discount factor δ fulfills the following conditions:

$$\begin{aligned} \delta &\geq \frac{n-1}{n+1} \quad \text{for } n \text{ odd}, \\ \delta &\geq \frac{n-1}{n} \quad \text{for } n \text{ even}. \end{aligned}$$

Thus, if $\delta \geq 1 - \frac{1}{n}$, the grand coalition is a WRP equilibrium in the complete network.

We conclude that whenever countries are homogeneous with respect to the number of neighbors in the network, global cooperation can be sustained as a WRP coalition. As derived in the comparative statics section, the more asymmetric the network becomes, the harder it is to restrain countries from renegotiation. For example, in the very asymmetric case of the star network, the grand coalition fails to be a WRP coalition.

Obviously, when the marginal local impact becomes negligible, the network structure, even if very asymmetric, becomes less significant, implying that global cooperation can be sustained in all networks.

Proposition 4. *Let $\delta \to 1$ and the local spillover be small enough. Then for every network there exists a WRP equilibrium supporting the global IEA.*

The result comes without proof since it follows immediately from Asheim and Holtsmark (2009), where $\gamma = 0$ is assumed and holds by continuity in γ.

[18] Note that in the proof of Proposition 3 we even show that it is possible to find an WRP equilibrium supporting the grand coalition such that none of the punishers wants to renegotiate. Such an equilibrium concept is more restrictive and therefore the existence result is even stronger. Further it prevents unreasonable equilibria to appear for instance by adding isolated countries to the punishment sets.

6 Social Benefits and Costs

6.1 Social Benefits

While we have shed some lights on the conditions for individual rational behavior with respect to membership in a coalition, it is important to know for e.g. policy implications what the collective or total welfare effect of an International Environmental Agreement is. We consider the utilitarian welfare composed of the sum of all countries' utilities, which is given by

$$\begin{aligned}\mathcal{W}(x(C)) = {} & n\Big(\frac{1}{2}(\gamma^2-\beta^2)\Big) - n\beta\sum_{i\in N}\bar{x}_i + \gamma\beta\Big(\sum_{i\in N}\eta_i + n^2\Big) + \beta^2 n^2 + \gamma^2\sum_{i\in N}\eta_i \\ & + \beta^2 k(k-1)\Big(n - \frac{1}{2}(k+1)\Big) + \beta\gamma\sum_{i\in N}k_i(k-1) + \gamma^2\sum_{i\in N}\sum_{m\in C}\bar{g}_{im}k_m \\ & + \beta\gamma n\sum_{m\in C}k_m - \frac{1}{2}\gamma\sum_{m\in C}k_m(2\beta k + \gamma k_m + 2\gamma) - \gamma\sum_{i\in N}\sum_{j\in N}\bar{g}_{ij}\bar{x}_j.\end{aligned}$$

Since in the global IEA all countries already maximize $\mathcal{W}$, it is immediate to see that the global IEA maximizes welfare. Further, because emission reduction always has a positive effect on all and the members of a coalition maximize the sum of utilities of their members, it is also quite immediate to see every IEA yields higher welfare than any of its subsets. Given an IEA C, the marginal effect of an additional member m on welfare can be calculated to be

$$\Delta(C\cup\{m\},C) = \beta^2 k\Big(2n - \frac{3}{2}(k+1)\Big) + \beta\gamma\Big(\sum_{i\notin C}k_i + k_m(2n-2-k)\Big) + \frac{\gamma^2}{2}k_m(k_m-1),$$

which is obviously positive. Thus, even though it might not be individually rational for some countries to join a coalition, the total welfare effect remains positive as the other countries' additional benefits outweigh the losses of that one individual country.

Further, it is easy to see that increasing spillovers, either through the relative effect β or γ, or the spillover structure (by adding links to the network) have negative effects on welfare.

6.2 Social Costs of Punishment

Besides the social benefits of a coalition there are also social costs whenever a country needs to be punished. Although this is off-equilibrium, one might ask what the welfare effect of punishment is and who should punish in cases when there is more than one possible punishment group that sustains global cooperation (see e.g., Example 2).

Therefore, assume a player j has deviated and the set of punishers P_j is called upon to punish. To study the effect of an additional punisher, denote by $\Delta\mathcal{W}\left(P_j, i\right)$ the marginal effect on welfare when player i joins the set of punishers P_j.

Lemma 1. *Suppose player j has deviated. Then,*

$$\Delta\mathcal{W}\left(P_j, i\right) = -\frac{1}{2}\Big(\beta(n-1) + \gamma\eta_i\Big)^2.$$

Now when allocating the set of punishers, the question arises who should punish; neighbors or non-neighbors? Recall that the marginal deterring effect of an additional punisher $i \in N$ on deviator $j \in N$ is given by

$$f_j(\delta, C, g, P_j \cup \{i\}) - f_j(\delta, C, g, P_j) = (\beta + \mathbb{1}_{N_j}(i)\gamma)(\beta(k-1) + \gamma k_i).$$

Then it is clear that in order to achieve an equal deterring effect, a non-neighbor $m \notin N_j$ must punish more, i.e. have more neighbors than a neighbor $i \in N_j$, i.e. $\eta_m > \eta_i$ which implies higher social costs. Hence, consider the case that two instead of one non-neighbor punishes.[19] Similarly to above, if we have $\eta_m > \eta_i$ for a $m \in P_j$, $m \notin N_j$, then the social cost of punishment will be larger when the non-neighbors punish. Instead, consider the case where both non-neighbors have smaller degree than a punishing neighbor, but together achieve the same deterring effect. The following result characterizes conditions on β and γ such that it is socially optimal to have a neighbor with higher degree punish.

Proposition 5. *Suppose that $\beta \leq (1+\sqrt{2})\gamma$. Then, punishment of a deviator by one of its neighbors is socially preferred to punishment by one or two non-neighbors such that the deterring effect is the same.*

Thus, if the global spillover effect β is not too large relative to the local spillover effect γ, it will be better in terms of welfare to have neighbors punish instead of non-neighbors since to achieve the same deterring effect, total punishment emission is higher when non-neighbors punish.

7 Extensions

7.1 Other Punishment Strategies

Besides the very specific penance punishment strategy we consider in our analysis above, there are of course numerous other ways to punish a possible deviator. Here, we want to discuss two possible variations of punishment strategies and their implications on the existence of WRP coalitions in the the repeated game.

[19] Of course, this may also have negative effects on the WRP condition since potentially two punishers' neighbors instead of one join the set of punishers. Here, however, we are only interested in the social cost of punishment.

7.1.1 Stronger Punishment

We have seen that in very asymmetric networks, such as the star, the grand coalition fails to be a WRP equilibrium with punishment levels $x_j^P = x_j^{NS}$. Let us now consider what happens if the punishment level that is emitted by the punishers is larger than their respective Nash output, i.e. what if $x_j^P > x_j^{NS}$ holds?

First, consider again the example of the star network.

Example 3. Let $\gamma = 1.5\beta$, $n = 5$ and $\delta \to 1$. For $p = 1$, the grand coalition can not be sustained as a WRP equilibrium but can we find a level $p^* < 1$ such that global cooperation is a WRP coalition? Suppose we want only three peripheral stars to punish the center node i, i.e. $|P_i| = 3$, and let us now determine the required punishment level p^* that yields a punishment strategy that sustains full cooperation as a WRP equilibrium. The center i has no incentive to unilaterally deviate from the signatory emission level if

$$\begin{aligned} (8) \quad &\Leftrightarrow \quad 3(5-p) + 1.5(3(7-2p)) + 2.25(3(2-p)) \geq 50 \\ &\Leftrightarrow \quad p^* \leq 0.53. \end{aligned} \tag{14}$$

is fulfilled. Furthermore, we have to ensure that no punisher $j \in P_i$ wants to unilaterally deviate from the punishment strategy. In the proof of Theorem 1 we used Lemma 2 to reduce the number of conditions for subgame perfection. As the Lemma generally only holds for $p \geq 1$, we can not directly transfer the conditions of Theorem 1 to this setting with a different punishment level. However, as long as $k_j > 0$ for all $j \in N$, the Lemma also holds for $p \geq 0$ and therefore if (14) is satisfied, full cooperation is an SGP equilibrium.

For the WRP conditions we have

$$\begin{aligned} (9) \quad &\Leftrightarrow \quad (5-p)(3-p) + 1.5((3-p)(1-p)+2) + \frac{2.25}{2}(1-p)^2 \leq 15.125 \\ &\Leftrightarrow \quad p^* \geq 0.6. \end{aligned}$$

Thus, $|P_i| = 3$ does also not yield a different result.

For $|P_i| = 2$ and $|P_i| = 1$, the condition for subgame perfection requires $p^* < 0$, thus $x_j^P > \bar{x}_j$ – *a contradiction*! Therefore, in this setting global cooperation fails to be a WRP equilibrium for any punishment level if only emitted for one period.

A more general statement, though, is not possible as the comparative static effects on the conditions for subgame perfection and weak renegotiation-proofness work in opposite directions for decreasing p. We therefore conclude that our restriction on Nash punishment levels is not too critical. Moreover, as mentioned before, it is obvious that even with these simple strategies the design of suitable punishment strategies is everything but straightforward as proposed in previous papers.

7.1.2 Longer Punishment

As a second variation, consider punishment strategies that punish a deviator for more than one period. More specifically, we change the simple strategy profile $\mathbf{s}$ to $\mathbf{s}^T$ such that we allow punishments over multiple but finite periods $T \in N$, i.e. the punishment path is given by

$$\mathbf{p_i^C}(T) = \Big\{ \underbrace{\Big(x^P_{P_i}, x^S_{C\setminus P_i}, x^{NS}_{N\setminus C}\Big), \ldots, \Big(x^P_{P_i}, x^S_{C\setminus P_i}, x^{NS}_{N\setminus C}\Big)}_{T \text{ periods}}, \Big(x^S_C, x^{NS}_{N\setminus C}\Big), \ldots \Big\}.$$

We assume that if during a punishment phase a new deviation occurs, either by the same or by another player, punishment switches to the beginning of the punishment path of that player. The conditions for subgame perfection and weak renegotiation-proofness then read as follows. First, there are no unilateral deviations from the equilibrium if

$$\sum_{t=1}^{T} \delta^t \Big(\pi_i(x^S_C) - \pi_i(x^S_i, x^P_{P_i}, x^S_{C\setminus P_i})\Big) \geq \pi_i(x^{NS}_i, x^S_{C\setminus\{i\}}) - \pi_i(x^S_C) \tag{15}$$

is satisfied for all $i \in N$. Furthermore, deviations from the punishment are deterred if

$$\delta^T \Big(\pi_j(x^S_C) - \pi_j(x^S_j, x^P_{P_j}, x^S_{C\setminus P_j})\Big) + \sum_{t=1}^{T-1} \delta^t \Big(\pi_j(x^P_{P_i}, x^S_{C\setminus P_i}) - \pi_j(x^S_j, x^P_{P_j}, x^S_{C\setminus P_j})\Big) \\ \geq \pi_j(x^S_j, x^P_{P_i\setminus\{j\}}, x^S_{C\setminus P_i}) - \pi_j(x^P_{P_i}, x^S_{C\setminus P_i}) \tag{16}$$

is fulfilled for all $j \in P_j$ and for all $i \in N$. Finally, for weak renegotiation-proofness, we need for all $i \in N$ at least one $j \in P_i$ such that

$$\sum_{t=0}^{T-1} \delta^t \Big(\pi_j(x^P_{P_i}, x^S_{C\setminus P_i}) - \pi_j(x^S_C)\Big) \geq 0$$

is satisfied, which is obviously equivalent to the original condition (9) of Theorem 1 with only one punishment period, i.e. $T = 1$. Consequently, the extension of the punishment period has no effect on the renegotiation incentives of the punishers.

Meanwhile, an increase in punishment periods does affect subgame perfection: the series on the left-hand side of (15) is equal to

$$\frac{\delta(1-\delta^T)}{1-\delta}\Big(\pi_i(x^S_C) - \pi_i(x^S_i, x^P_{P_i}, x^S_{C\setminus P_i})\Big),$$

which obviously increases for larger T but is bounded from above by

$$\frac{\delta}{1-\delta}\Big(\pi_i(x^S_C) - \pi_i(x^S_i, x^P_{P_i}, x^S_{C\setminus P_i})\Big).$$

We receive that independent of the parameters, an extension of the punishment yields that fewer punishers are sufficient to deter a country from deviating from the signatory strategy. In line with the folk theorem we can conclude that in any given network g, if players are patient enough, i.e. if δ is sufficiently large, we can always find a duration of punishments T such that for (15) to be fulfilled, a single punisher is sufficient. Also, this punisher does not need to be a neighbor.

Additionally, for small enough P_i, the left-hand side of (16) is always positive such that we can have that for a large enough T, (16) is always satisfied, too. Thus, we can achieve full cooperation as an SGP coalition. As WRP is not affected and for $|P_i| = 1$ it is always satisfied, we can conclude without proof the following theorem:

Proposition 6. *For δ sufficiently large, for every network g there exists a duration of punishments T such that the simple strategy profile $\mathbf{s}^T$ sustains full cooperation as a WRP coalition.*

Note, however, that this result requires very long punishment and hence lots of pollution. Such a threat might not be credible, if e.g. the negative consequences of accumulated pollution are increasing in the amount of pollution. We therefore conclude again that our restriction on the simple strategy profile is not too restrictive and already offers several interesting insights into the structure of the model.

Example 4. Let $\gamma = 1.5\beta$, $n = 5$ and $\delta \to 1$. We have seen that for $p = 1$, the grand coalition can not be sustained as a WRP equilibrium and also harsher punishment has not changed this result due to the large asymmetry between the center and peripheral nodes. Now for $p = 1$, we can find $T > 1$ such that global cooperation is a WRP coalition:

Let $T = 2$. Then, as WRP remains unchanged, for the center node i the punishment group must not be larger than 3. Condition (15) for the center node yields $P_i \geq 2$ so it remains to check condition (16). Let us choose $P_i = 3$. For any peripheral node j, condition (15) yields that the center node is sufficient to punish, i.e. $|P_j| = 1$. Then, with these punishment sets given, (16) is satisfied for the center and also for the peripheral nodes. Thus, when the punishment phase is extended to two periods, full cooperation can be sustained as a WRP equilibrium in this network.

7.2 Directed Networks

In our model we have restricted ourselves to the consideration of undirected networks, i.e. local spillovers are assumed to be bidirectional. Of course, this may often not be the case. For example, the direction of the winds play an important role for the effects of air pollution and the direction a river flows influences the pollution effects along the stream.

Intuitively, to analyze for directed networks in our model, we only have to slightly modify our notation and denote by ij the link in the network that represents a spillover from player i to player j. Denote by $\overrightarrow{N_i}$ the set of outgoing links, i.e. players that player i

is connected to and by $\overleftarrow{N_i}$ the set of incoming links, i.e. players that are connected to i. We also alter the definition for η_i and k_i accordingly. When choosing the optimal signatory output emission x_i^S, country i considers its own contribution to the global spillover and its local impact on the countries it is directly linked to. That is, the signatory's output is given by $x_i^S = \bar{x}_i - \beta k - \gamma(\overrightarrow{k_i} + 1)$.

For stability, we now have to distinguish between the incoming and the outgoing links. For example, the more outgoing links, the higher the incentive to deviate. The more incoming links, the less punishers are necessary to credibly deter a deviation. Therefore, the asymmetry in a network can now be two-folded and prevent the global IEA to be sustainable as WRP equilibrium. In order to extend our existence result of Proposition 3 to directed networks, we need to adapt the definition of regularity: we call a directed network a *regular* network, if for all players the number of incoming links and the number of outgoing links are equal, i.e. if $\overleftarrow{\eta_i} = \overrightarrow{\eta_i} = \eta$ for all $i \in N$. Then, the proof of Proposition 3 is analogous and we receive that also in directed networks, global cooperation can be sustained as WRP equilibrium if there is no asymmetry with respect to the network structure.

7.3 Formation of an IEA

So far we have restricted the analysis to the question of stability of IEAs by means of WRP equilibria without modeling the formation of an IEA. We briefly and informally outline here, how such an IEA could come into place. First, of course, we could always imagine a climate conference where the local spillover structures are taken into account. Since there are potentially many equilibria of the repeated game even for a given coalition, it is difficult to model the strategies used by the countries to select among the WRP equilibria of repeated game.

In fact we could also imagine that a small subset of all countries (e.g. the US and Canada, or the countries within the EU) start out with a coalition to obtain a critical mass and then approach countries outside the coalition, particularly those exposed to local spillovers of the coalition by offering those countries to reduce emission if they themselves do so. In other words, they threaten punishment by non-implementation of an IEA (which is equivalent to business as usual) if other countries do not reduce themselves. Thus, given a coalition C_1, rank all other countries $i \in N \setminus C_1$ by the ration $\frac{|N_i \cap C_1|}{|N_i|}$. Coalition C_1 then agrees to the terms of an IEA conditional on additional countries joining. Those with large ratio $\frac{|N_i \cap C_1|}{|N_i|}$ are the ones that are most likely to join C_1 since for them the SGP condition (8) is easiest to be satisfied by the threat of non-implementation of signatory strategies of C_1. Thus i_1 would join C_1 if there exists a punishment set $P_{i_1} \subset C_1$ such that the conditions of Theorem 1 are satisfied. After acquiring the highest ranked country i_1 to C_1, for $C_2 = C_1 \cup \{i_1\}$ repeat the procedure for C_2 etc.

Since small coalitions are easy to sustain even without threats of future punishment (cf. Section 3.2), it may be the case that such a procedure actually leads to the implementation of a global IEA (if stable for the given spillover structure). In this way, an initially small IEA may spread to a large IEA, i.e. from an initially local IEA can emerge a global IEA (cf. also Section 3.4.2 in Currarini et al., 2014).

8 Conclusion

We have merged local and global pollution spillovers into one model by introducing a network structure. In the single-stage game model this has not dramatically changed the well-known results. Caused by the free-rider incentives, self-enforcing cooperation is only achieved for very few countries and does thus not significantly contribute to the reduction of global pollution.

In the repeated game, using a specific punishment strategy, weakly renegotiation-proof agreements can be achieved via the threat of punishments. If the punishing countries suffer too much from punishment themselves, they may want to renegotiate. To account for this, we characterize an individual group of punishing countries for each coalition member and therefore decrease the incentives to renegotiate. However, when the network is very asymmetric, as for example in the star network, full cooperation may not be a WRP coalition. In turn, when players are symmetric with respect to their spillover impacts and sufficiently patient, in regular networks the grand coalition can be sustained as a weakly renegotiation-proof equilibrium.

Finally we analyzed welfare implications of the network structure. More links in the network have a negative impact on global welfare as the local spillover effects outweigh the higher efforts by signatories that internalize the additional externality.

Due to the generality of our approach, our model can serve as benchmark model which should be extended and refined in the future. Yet we can already see that, along the lines of Bollen et al. (2009), a pollution policy that takes account of the effects of both global and local (air) pollution can help sustain global cooperation and ultimately increase global welfare. By taking into account the local spillover structure, punishment mechanisms can be designed more appropriately and therefore help deter countries from free-riding without making the agreement vulnerable to renegotiation. In this way, a few countries (e.g. US and Canada or EU countries) who initially agree to certain terms of reduction conditional on others joining them, may achieve global cooperation by particularly taking the spillover structure into account. Our model is also not limited to the application in the strive for joint emission reduction. It can easily be adapted to other problems in the provision of public goods.

In a next step, some simplifications we have taken may be relaxed. Of course, further heterogeneities imply less analytical tractability but as it is frequently done in the IEA literature, simulations could be considered to compare the outcomes of our model to other existing ones. Furthermore, other punishment strategies and the possibility of multiple coalitions may be worth studying, too.

Regarding multiple coalitions, as long as we strive for the socially optimal outcome of global cooperation, there is no need for more than a single agreement. However, whenever global cooperation can not be sustained as a self-enforcing equilibrium, one could study what happens if multiple coalitions would form. In the case of linear costs of pollution, one could reach an outcome where every country is a signatory in a (possibly only very small coalition). Then, individual contributions to abatement may not substantially improve the business as usual outcome as all countries only account for very few externalities of their coalition members.

Also, there are several ways the model proposed in this paper could be extended. As noted in Currarini et al. (2014), there is a large potential for network economics to be applied in environmental economics. In the following, we want to discuss several of the aspects that could emerge from our model.

First, the extension of the model to incorporate transfers and side-payments is natural. One could then interpret the underlying network structure, i.e. the links between countries, also as established ways of communication or negotiation through which countries can offer side-payments to incentivize non-cooperators to join the coalition.

Second, the network could also represent a different underlying structure, for instance an established trade structure that enables countries to link pollution and trade strategies (*issue-linkage*). Trade and other related issues have been subject of the environmental economics research and our model could generate new results using techniques from network economics.

Third, while reduction of emissions is one way to contribute to the global effort of fighting climate change, investments in R&D is another possibility to mitigate pollution. And as it is standard in the (network) literature, spillovers from R&D play an important role in the decision of optimal investments. Thus, bringing together the literature of R&D spillovers and the mitigation of pollution through an IEA is another possible extension of our model.

Fourth, there already exist some models that study local and regional agreements that may lead to global cooperation. Methods from Evolutionary Game Theory have been used to study whether or not local agreements may facilitate the formation of global cooperation.[20] By applying results from opinion formation in a network, our model may serve as an approach to better understand the chances of such a formation process. In our benchmark model we only consider the formation of a single IEA, but the extension to multiple coalitions should be natural and thus offer a promising area of future research.

[20] Regional agreements and initiatives have been formed to tackle the problem of regional pollution effects (visit the Global Atmospheric Pollution Forum online for a list of regional initiatives worldwide). One example is the "Climate and Clean Air Coalition" that strives for a reduction of short-lived air pollutants and has been gaining influence over the past years.

Appendix

A Proofs

Proof of Proposition 1. When all countries cooperate on abatement, the profit of free-riding on these efforts is given by

$$\begin{aligned}\pi_i(N \setminus \{i\}) = &-\frac{1}{2}(\beta+\gamma)^2 - (\beta+\gamma)(\bar{x}_i - \beta - \gamma) - \beta\sum_{j\neq i}\big(\bar{x}_j - \beta(n-1) \\ &-\gamma(\eta_j + 1 - g_{ij})\big) - \gamma\sum_{j\neq i} g_{ij}\Big(\bar{x}_j - \beta(n-1) - \gamma(\eta_j + 1 - g_{ij})\Big).\end{aligned}$$

Thus, we have for all $i \in N$

$$\begin{aligned}\pi_i(N) - \pi_i(N \setminus \{i\}) =& -\frac{1}{2}\Big(\big(\beta n + \gamma(\eta_i+1)\big)^2 - (\beta+\gamma)^2\Big) - (\beta+\gamma)\left(-\beta n - \gamma(\eta_i+1) + \beta + \gamma\right) \\ &- \sum_{j\neq i}(\beta + \gamma g_{ij})\left(-\beta n - \gamma(\eta_j+1) + \beta(n-1) + \gamma(\eta_+1 - g_{ij})\right) \\ =& -\frac{1}{2}\Big(\big(\beta n + \gamma(\eta_i+1)\big)^2 - (\beta+\gamma)^2\Big) + (\beta+\gamma)\left(\beta(n-1) + \gamma\eta_i\right) + \sum_{j\neq i}(\beta+\gamma g_{ij})^2 \\ =& \,\beta^2\Big(2n - \frac{3}{2} - \frac{1}{2}n^2\Big) + \beta\gamma\eta_i\Big(3-n\Big) + \gamma^2\eta_i\Big(1 - \frac{1}{2}\eta_i\Big),\end{aligned}$$

which is clearly negative for $n > 3$ and $\eta_i \neq 1$.

For $\eta_i = 1$, the difference above reduces to

$$\beta^2\Big(2n - \frac{3}{2} - \frac{1}{2}n^2\Big) + \beta\gamma\Big(3-n\Big) + \frac{\gamma^2}{2},$$

which is negative if and only if $\gamma < \beta\left(n - 3 + \sqrt{(n-3)(2n-4)}\right)$ holds. □

Proof of Theorem 1. In order for **s** to be a subgame perfect equilibrium (SGP), there are two conditions that need to be fulfilled for all signatories $i \in C$:

(i) No signatory $i \in C$ has an incentive to deviate from x_i^S

(ii) Given country $i \in C$ deviates, no punishing country $j \in P_i$ has an incentive to not punish i

For condition (i) to be satisfied, we have to derive conditions such that the following holds for all $i \in C$:

$$\begin{aligned} & \pi_i(x_C^S) + \delta\pi_i(x_C^S) \geq \pi_i(x_i^{NS}, x_{C\setminus\{i\}}^S) + \delta\pi_i(x_i^S, x_{C\setminus P_i}^S, x_{P_i}^P) \\ \Leftrightarrow \quad & \delta\left(\pi_i(x_C^S) - \pi_i(x_i^S, x_{C\setminus P_i}^S, x_{P_i}^P)\right) \geq \pi_i(x_i^{NS}, x_{C\setminus\{i\}}^S) - \pi_i(x_C^S) \end{aligned} \tag{17}$$

If all signatories $i \in C$ play the signatory strategy x_i^S as agreed upon, the discounted payoff for i is

$$\begin{aligned}
\pi_i(x_C^S) = &- \frac{1}{2}\big(\beta k + \gamma(k_i+1)\big)^2 - \beta \sum_{m\in C} \big(\bar{x}_m - \beta k - \gamma(k_m+1)\big) - \beta \sum_{l \notin C} (\bar{x}_l - \beta - \gamma) \\
&- \gamma \sum_{m \in C} \bar{g}_{im}\big(\bar{x}_m - \beta k - \gamma(k_m+1)\big) - \gamma \sum_{l\notin C} \bar{g}_{il}(\bar{x}_l - \beta - \gamma).
\end{aligned}$$

Consider now a situation when country $i \in C$ deviates from x_i^S in period t. Then, by $\mathbf{s}$, in the next period $t+1$ we have

$$x_j(C, t+1) = \begin{cases} \bar{x}_j - \beta k - \gamma(k_j+1), & \text{if } j = i \\ \bar{x}_j - p(\beta+\gamma), & \text{if } j \in P_i \cup (N \setminus C) \\ \bar{x}_j - \beta k - \gamma(k_j+1), & \text{if } j \in C \setminus P_i \end{cases}.$$

This yields the stage payoff

$$\begin{aligned}
\pi_i(x_i^S, x_{C\setminus P_i}^S, x_{P_i}^P) = &- \frac{1}{2}\big(\beta k + \gamma(k_i+1)\big)^2 - \beta \sum_{m \in C\setminus P_i} \big(\bar{x}_m - \beta k - \gamma(k_m+1)\big) \\
&- \beta \sum_{l\in P_i} \big(\bar{x}_l - p(\beta+\gamma)\big) - \beta \sum_{l\notin C} (\bar{x}_l - \beta - \gamma) - \gamma \sum_{l \notin C} \bar{g}_{il}(\bar{x}_l - \beta - \gamma) \\
&- \gamma \sum_{m\in C\setminus P_i} \bar{g}_{im}\big(\bar{x}_m - \beta k - \gamma(k_m+1)\big) - \gamma \sum_{l \in P_i} \bar{g}_{il}\big(\bar{x}_l - p(\beta+\gamma)\big),
\end{aligned}$$

and we receive for the payoff loss from a one-shot deviation

$$\begin{aligned}
\pi_i(x_C^S) - \pi_i(x_i^S, x_{C\setminus P_i}^S, x_{P_i}^P) &= -\beta \sum_{m\in C} \big(\bar{x}_m - \beta k - \gamma(k_m+1)\big) \\
&- \gamma \sum_{m\in C} \bar{g}_{im}\big(\bar{x}_m - \beta k - \gamma(k_m+1)\big) + \beta \sum_{m\in C\setminus P_i} \big(\bar{x}_m - \beta k - \gamma(k_m+1)\big) \\
&+ \beta \sum_{l\in P_i} \big(\bar{x}_l - p(\beta+\gamma)\big) + \gamma \sum_{m \in C\setminus P_i} \bar{g}_{im}(\bar{x}_m - \beta k - \gamma(k_m+1)) \\
&+ \gamma \sum_{l\in P_i} \bar{g}_{il}(\bar{x}_l - p(\beta+\gamma)) \\
&= -\beta \sum_{l\in P_i} [(\bar{x}_l - \beta k - \gamma(k_l+1)) - (\bar{x}_l - p(\beta+\gamma))] \\
&- \gamma \sum_{l\in P_i} g_{il}[(\bar{x}_l - \beta k - \gamma(k_l+1)) - (\bar{x}_l - p(\beta+\gamma))] \\
&= \beta \sum_{l\in P_i} [\beta(k-p) + \gamma(k_l+1-p)] + \gamma \sum_{l\in P_i} g_{il}[\beta(k-p) + \gamma(k_l+1-p)]. \qquad (18)
\end{aligned}$$

Furthermore we have for the short-term payoff gain from a one-shot deviation

$$\begin{aligned}\pi_i(x_i^{NS}, x^S_{C\backslash\{i\}}) - \pi_i(x^S_C) &= \frac{1}{2}\Big(\beta k + \gamma(k_i+1)\Big)^2 - \frac{1}{2}\Big(\beta+\gamma\Big)^2 - \Big(\beta+\gamma\Big)\Big(\beta(k-1)+\gamma k_i\Big)\\ &= \frac{\beta^2}{2}(k-1)^2 + \beta\gamma(k_i(k-1)) + \frac{\gamma}{2}k_i^2\\ &= \frac{1}{2}\left(\beta(k-1)+\gamma k_i\right)^2 \geq 0. \qquad (19)\end{aligned}$$

Multiplying with δ and rewriting equation (18), then substracting (19), we obtain that condition (i) is satisfied if (8) holds.

Let us now consider condition (ii). Suppose country i deviated in period $t-1$. In order to ensure that all $j \in P_i$ actually punish the deviator, the following condition has to hold for all $j \in P_i$:

$$\begin{aligned}\pi_j(x^P_{P_i}, x^S_{C\backslash P_i}) + \delta\pi_j(x^S_C) &\geq \pi_j(x^S_j, x^P_{P_i\backslash\{j\}}, x^S_{C\backslash P_i}) + \delta\pi_j(x^S_j, x^S_{C\backslash P_j}, x^P_{P_j})\\ \delta\Big(\pi_j(x^S_C) - \pi_j(x^S_j, x^S_{C\backslash P_j}, x^P_{P_j})\Big) &\geq \pi_j(x^S_j, x^P_{P_i\backslash\{j\}}, x^S_{C\backslash P_i}) - \pi_j(x^P_{P_i}, x^S_{C\backslash P_i}) \qquad (20)\end{aligned}$$

For the single-stage payoffs we have

$$\begin{aligned}\pi_j(x^S_j, x^P_{P_i\backslash\{j\}}, x^S_{C\backslash P_i}) = &-\frac{1}{2}\left(\beta k + \gamma(k_j+1)\right)^2 - \beta\sum_{m\in P_i\backslash\{j\}}\left(\bar{x}_m - p(\beta+\gamma)\right) - \beta\sum_{m\notin C}(\bar{x}_m - \beta - \gamma)\\ &-\beta\sum_{l\in C\backslash(P_i\backslash\{j\})}\left(\bar{x}_l - \beta k - \gamma(k_l+1)\right) - \gamma\sum_{m\in P_i\backslash\{j\}}\bar{g}_{jm}\left(\bar{x}_m - p(\beta+\gamma)\right)\\ &-\gamma\sum_{m\notin C}\bar{g}_{jm}(\bar{x}_m - \beta-\gamma) - \gamma\sum_{l\in C\backslash(P_i\backslash\{j\})}\bar{g}_{jl}(\bar{x}_l - \beta k - \gamma(k_l+1))\end{aligned}$$

and

$$\begin{aligned}\pi_j(x^P_{P_i}, x^S_{C\backslash P_i}) = &-\frac{1}{2}\left(p(\beta+\gamma)\right)^2 - \beta\sum_{m\in P_i}\left(\bar{x}_m - p(\beta+\gamma)\right) - \beta\sum_{m\notin C}(\bar{x}_m - \beta-\gamma)\\ &-\beta\sum_{l\in C\backslash P_i}\left(\bar{x}_l - \beta k - \gamma(k_l+1)\right) - \gamma\sum_{m\in P_i}\bar{g}_{jm}\left(\bar{x}_m - p(\beta+\gamma)\right)\\ &-\gamma\sum_{m\notin C}\bar{g}_{jm}(\bar{x}_m - \beta-\gamma) - \gamma\sum_{l\in C\backslash P_i}\bar{g}_{jl}\left(\bar{x}_l - \beta k - \gamma(k_l+1)\right).\end{aligned}$$

We will show that (17) already implies (20). To prove this, we need the following Lemma.

Lemma 2. *For all β, γ, k, k_j and $p \geq 1$ it always holds*

$$\begin{aligned}&-\frac{1}{2}(\beta k + \gamma(k_j+1))^2 + \frac{1}{2}(p(\beta+\gamma))^2 + (\beta+\gamma)(\beta(k-p) + \gamma(k_j+1-p))\\ &\qquad\leq \frac{1}{2}(\beta k + \gamma(k_j+1))^2 - \frac{1}{2}(\beta+\gamma)^2 - (\beta+\gamma)(\beta(k-1)+\gamma k_j). \qquad (21)\end{aligned}$$

Proof. As $x_j^P \geq x_j^S$, we have $p(\beta + \gamma) \leq \beta k + \gamma(k_j + 1)$ for all j in P_i and thus we have

$$\begin{aligned}
0 &\leq \big(\beta(k-1) + \gamma k_j\big)^2 - \frac{1}{2}\left((p-1)(\beta+\gamma)\right)^2 \\
&= \beta^2\Big((k-1)^2 - \frac{1}{2}(p-1)^2\Big) + \gamma^2\Big(k_j^2 - \frac{1}{2}(p-1)^2\Big) + \beta\gamma\Big(2k_j(k-1) + (1-p)^2\Big) \\
&= \beta^2\Big(k^2 - \frac{1}{2}(1+p^2) - (2k-p-1)\Big) + \gamma^2\Big((k_j+1)^2 - \frac{1}{2}(1+p^2) - (2k_j+1-p)\Big) \\
&\quad + \beta\gamma\Big(2k(k_j+1) - p^2 - (2k-p-1+2k_j+1-p)\Big) \\
&= \Big(\beta k + \gamma(k_j+1)\Big)^2 - \frac{1}{2}\Big((\beta+\gamma)^2 + (p(\beta+\gamma))^2\Big) - (\beta+\gamma)\left(\beta(2k-p-1) + \gamma(2k_j+1-p)\right),
\end{aligned}$$

which is nothing else but (21) and proves the lemma. □

We can now rewrite the left-hand side of (20) and receive

$$\begin{aligned}
&\pi_j(x_j^S, x_{P_i\setminus\{j\}}^P, x_{C\setminus P_i}^S) - \pi_j(x_{P_i}^P, x_{C\setminus P_i}^S) \\
&= -\frac{1}{2}\left(\beta k + \gamma(k_j+1)\right)^2 + \frac{1}{2}\left(p(\beta+\gamma)\right)^2 - \beta\left(\bar{x}_j - \beta k - \gamma(k_j+1)\right) + \beta\left(\bar{x}_j - p(\beta+\gamma)\right) \\
&\quad - \gamma\left(\bar{x}_j - \beta k - \gamma(k_j+1)\right) + \gamma\left(\bar{x}_j - p(\beta+\gamma)\right) \\
&= -\frac{1}{2}\left(\beta k + \gamma(k_j+1)\right)^2 + \frac{1}{2}\left(p(\beta+\gamma)\right)^2 + (\beta+\gamma)\left(\beta(k-p) + \gamma(k_j+1-p)\right) \\
&\leq \frac{1}{2}\left(\beta k + \gamma(k_j+1)\right)^2 - \frac{1}{2}(\beta+\gamma)^2 - (\beta+\gamma)\left(\beta(k-1) + \gamma k_j\right) \\
&= \frac{1}{2}\left(\beta(k-1) + \gamma k_j\right)^2 \\
&= \pi_i(x_i^{NS}, x_{C\setminus\{i\}}^S) - \pi_i(x_C^S).
\end{aligned}$$

Thus, whenever (17) is satisfied, (20) has no bite and **s** therefore constitutes an SGP equilibrium if and only if (17) holds.

Let us now turn to the condition of weak renegotiation-proofness. As given in Definition 1, a subgame perfect equilibrium **s** is weakly renegotiation-proof (WRP) if there do not exist two continuation equilibria such that all players strictly prefer the one to the other. That is, we have to derive conditions such that all punishing countries $j \in P_i$ will actually punish instead of ignoring the deviation and continuing with another equilibrium path, e.g. essentially renegotiating to playing cooperate again.

For any period t, there are $k+1$ possible continuation equilibria that implement either the agreement path $\mathbf{a^C}$ or the punishment path $\mathbf{p_j^C}$ for any signatory $j \in C$.

Assume that the strategy profile **s** is an SGP, thus condition (8) is satisfied. In accordance with the definition, for weak renegotiation-proofness we now need to consider all continuation equilibria and the respective incentives of each player.

Obviously, all signatories prefer the agreement continuation equilibrium to the one generated from their respective punishment path $\mathbf{p_i^C}$, i.e. $\pi_i(x_C^S) > \pi_i(x_{C\setminus P_i}^S, x_{P_i}^P)\ \forall\, i \in C$.

Thus, any country that is punished would obviously not block a renegotiation to the agreement path.

All non-signatories $j \notin C$ will continue to free-ride on the others' efforts in any continuation equilibrium. They will not block a renegotiation either. Also, all signatories $j \in C \setminus P_i$ that do not punish prefer the equilibrium path with payoffs $\pi_j(x_C^S)$ to a continuation equilibrium from following the punishment path $\mathbf{p_i^C}$.

Thus, it remains to check the incentives of the punishers. If $\pi_j(x_C^S) > \pi_j(x_{P_i}^P, x_{C\setminus P_i}^S)$ holds for all $j \in P_i$, all punishing countries prefer the continuation equilibrium when no punishing is carried out to the one where i deviated. Thus, all players strictly prefer the agreement path to the punishment path and therefore $\mathbf{s}$ would not be weakly renegotiation-proof. Therefore, if there exists a punisher $j \in P_i$ such that $\pi_j(x_C^S) \leq \pi_j(x_{P_i}^P, x_{C\setminus P_i}^S)$, renegotiation would be blocked and $\mathbf{s}$ would indeed be a WRP equilibrium.

Hence, for $\mathbf{s}$ to be a WRP equilibrium the following condition needs to be satisfied for at least one $j \in P_i$:

$$\pi_j(x_{P_i}^P, x_{C\setminus P_i}^S) - \pi_j(x_C^S) \geq 0. \tag{22}$$

We have

$$\begin{aligned}
\pi_j(x_{P_i}^P, x_{C\setminus P_i}^S) - \pi_j(x_C^S) = & -\frac{1}{2}\left(p(\beta+\gamma)\right)^2 + \frac{1}{2}\left(\beta k + \gamma(k_j+1)\right)^2 \\
& - \beta \sum_{l \in P_i} \left(\bar{x}_l - p(\beta+\gamma)\right) + \beta \sum_{m \in P_i} \left(\bar{x}_m - \beta k - \gamma(k_m+1)\right) \\
& - \gamma \sum_{l \in P_i} \bar{g}_{jl} \left(\bar{x}_l - p(\beta+\gamma)\right) + \gamma \sum_{m \in P_i} \bar{g}_{jm} \left(\bar{x}_m - \beta k - \gamma(k_m+1)\right) \\
= & -\frac{1}{2}\left(\left(p(\beta+\gamma)\right)^2 - \left(\beta k + \gamma(k_j+1)\right)^2\right) - \beta \sum_{m \in P_i} \left(\beta(k-p) + \gamma(k_m+1-p\right) \\
& - \gamma \sum_{m \in P_i} \bar{g}_{jm} \left(\beta(k-p) + \gamma(k_m+1-p)\right)
\end{aligned}$$

and thus (22) is equivalent to

$$\begin{aligned}
\frac{\beta^2}{2}\left((k-p)(k+p-2|P_i|)\right) + \beta\gamma \left(kk_j - |P_i \cap N_j|(k-p) - \sum_{m \in P_i} (k_m+1-p) + p(1-p) \right) \\
+ \frac{\gamma^2}{2} \left(k_j^2 - 2\sum_{m \in P_i} g_{jm}(k_m+1-p) - 1 + p(2-p) \right) \geq 0,
\end{aligned} \tag{23}$$

which we can rewrite such that we obtain (9).

Concluding, if the strategy $\mathbf{s}$ satisfies (9) for all $i \in C$, i.e. is subgame perfect, and additionally is such that for any $i \in C$ there is a punishment set P_i such that there exists at least one $j \in P_i$ that satisfies condition (9), $\mathbf{s}$ is weakly renegotiation-proof. □

Proof of Corollary 1. For $\gamma = 0$ we have that the strategy $\mathbf{s}$ that implements coalition C is a subgame perfect equilibrium if and only if for all $i \in C$

$$\delta\beta^2|P_i|(k-1) \geq \frac{1}{2}\left(\beta(k-1)\right)^2$$

holds. Furthermore, it is weakly renegotiation-proof if and only if for all $i \in C$

$$\beta^2(k-1)(|P_i|-1) \leq \frac{1}{2}\left(\beta(k-1)\right)^2.$$

For $k \geq 2$ This gives

$$\begin{aligned} & |P_i| \geq \frac{k-1}{2\delta} \quad \wedge \quad |P_i| \leq \frac{k+1}{2} \\ \Leftrightarrow \quad & \frac{1}{2\delta}(k-1) \leq |P_i| \leq \frac{1}{2}(k+1). \end{aligned}$$

□

Proof of Proposition 2. As we have already seen in Example 1, global cooperation can not be supported as an SGP equilibrium in the star network if $\gamma > \beta\frac{n-1}{n-3}$ holds.

Let us now suppose $\gamma \leq \beta\frac{n-1}{n-3}$ and consider again the center i of the star network. Then, we can find a punishment group P_i such that full cooperation can be sustained as an SGP coalition, that is we can find P_i such that $|P_i| \geq \frac{(n-1)^2(\beta+\gamma)}{2(\beta(n-1)+\gamma)}$ is satisfied (compare condition (8) of Theorem 1). Clearly, whenever this lower bound is larger than the upper bound from the WRP condition (9), full cooperation can not be a WRP coalition. We have

$$\begin{aligned} & \frac{(n-1)^2(\beta+\gamma)}{2(\beta(n-1)+\gamma)} > \frac{\beta(n-1)+\gamma}{2\beta} \\ \Leftrightarrow \quad & n > n_1 := 2 + \frac{\beta}{\gamma}\left(1+\sqrt{1+2\frac{\gamma}{\beta}+3(\frac{\gamma}{\beta})^2+(\frac{\gamma}{\beta})^3}\right). \end{aligned} \tag{24}$$

Hence, given parameters γ and β, for n large enough (24) is always satisfied and global cooperation fails to be a WRP coalition. □

Proof of Proposition 3. Denote by Ψ_i the set of permutations of players $\psi_i : N \setminus \{i\} \to N \setminus \{i\}$ such that $\psi_i(j) < \psi_i(m)$ for all $j \in N_i$, $m \notin N_i$. Further, given a permutation $\psi_i \in \Psi_i$, let $M_{\psi_i}(\nu)$ denote the first ν elements of the permutation, i.e. $M_{\psi_i}(\nu) := \{\psi_i(1), \ldots, \psi_i(\nu)\}$. This defines a possible punishing set P_i.

Let $\nu^* := \arg\min_{1\leq\nu\leq n-1}\{P_i = M_{\psi_i}(\nu)$ satisfies (12)$\}$ be the lowest integer such that the set $M_{\psi_i}(\nu^*)$ deters a signatory from deviating.

First suppose that $\nu^* \leq \eta$. Then by construction we have $M_{\psi_i}(\nu^*) \cap N_i = M_{\psi_i}(\nu^*)$ and for any $j \in M_{\psi_i}(\nu^*)$, j is also in N_i. Thus, $j \notin M_{\psi_i}(\nu^*) \cap N_j$ and therefore we get for any $\psi_i \in \Psi_i$ that

$$|M_{\psi_i}(\nu^*) \cap N_j| \leq |M_{\psi_i}(\nu^*) \cap N_i| - 1 \tag{25}$$

holds for all $j \in M_{\psi_i}(\nu^*)$.

Hence, the following holds for all $\psi_i \in \Psi_i$:

$$|M_{\psi_i}(\nu^*) \cap N_j|\gamma + (|M_{\psi_i}(\nu^*)| - 1)\beta \leq (|M_{\psi_i}(\nu^*) \cap N_i| - 1)\gamma + (|M_{\psi_i}(\nu^*)| - 1)\beta. \tag{26}$$

Suppose now the opposite, i.e. $\nu^* > \eta$. We show that there still exists a permutation $\psi_i^* \in \Psi_i$ such that (26) holds for all $j \in M_{\psi_i^*}(\nu^*)$.

From (12) we get that $\nu^* \geq \frac{1}{2}(n-1) - \frac{1}{2}\frac{\gamma}{\beta}\eta$ since for $P_i = M_{\psi_i}(\nu^*)$ we have that $M_{\psi_i}(\nu^*) \cap N_i = N_i$. Moreover, because of minimality and ν^* being an integer, we have

$$\nu^* = \left\lceil \frac{1}{2}(n-1) - \frac{1}{2}\frac{\gamma}{\beta}\eta \right\rceil . \tag{27}$$

For all neighbors $j \in M_{\psi_i}(\nu^*) \cap N_i$ of the deviator i, (25) still holds and there is nothing to show.

From (27) we receive that additional to the neighbors of i there are $\nu^* - \eta = \left\lceil \frac{1}{2}(n-1) - \frac{\gamma}{2\beta}\eta \right\rceil - \eta$ non-neighbors in the punishing set $M_{\psi_i}(\nu^*)$.

Denote by $\tilde{\psi_i} \in \Psi_i$ the permutation which minimizes the number of those non-neighbors $j \in M_{\psi_i}(\nu^*) \setminus N_i$ that have all their links within the set $M_{\psi_i}(\nu^*)$, i.e. such that $\eta_j(g|_{M_{\psi_i}(\nu^*)}) = \eta$ holds.[21] Suppose that this number is different from zero, i.e. at least one country in $M_{\psi_i}(\nu^*) \setminus N_i$ has all neighbors in $M_{\psi_i}(\nu^*)$. Then, by (27), from the set $M_{\tilde{\psi}_i}(\nu^*)$ there are at most $\eta \left\lceil \left(\frac{1}{2}(n-1) - \frac{1}{2}\frac{\gamma}{\beta}\eta\right) \right\rceil - 2\eta$ links into the set $N \setminus \left\{ M_{\tilde{\psi}_i}(\nu^*) \cup \{i\} \right\}$.

As $|N \setminus \left\{ M_{\tilde{\psi}_i}(x^*) \cup \{i\} \right\}| = n - 1 - \nu^*$, we have that the sum of degrees of members of the set $N \setminus \left\{ M_{\tilde{\psi}_i}(x^*) \cup \{i\} \right\}$ satisfies

$$\begin{aligned}
\eta \left| N \setminus \left\{ M_{\tilde{\psi}_i}(\nu^*) \cup \{i\} \right\} \right| &= \eta \left(n - 1 - \left\lceil \left(\frac{1}{2}(n-1) - \frac{1}{2}\frac{\gamma}{\beta}\eta \right) \right\rceil \right) \\
&= \eta \left\lceil \left(\frac{1}{2}(n-1) + \frac{1}{2}\frac{\gamma}{\beta}\eta \right) \right\rceil > \eta \left\lceil \left(\frac{1}{2}(n-1) - \frac{1}{2}\frac{\gamma}{\beta}\eta \right) \right\rceil - 2\eta.
\end{aligned}$$

[21] Note that all permutations $\psi_i \in \Psi_i = \{\psi_i : N \setminus \{i\} \to N \setminus \{i\} \text{ s.t. } \psi_i(j) < \psi_i(m) \; \forall j \in N_i, m \notin N_i\}$, deliver the same ν^* due to regularity of the network.

Thus, the number of all links of members of the set $N \setminus \left\{M_{\tilde{\psi}_i}(\nu^*) \cup \{i\}\right\}$ exceeds the maximum amount of links coming into the set from its complement, meaning that there has to exist a link lm between members l, m of the set $N \setminus \left\{M_{\tilde{\psi}_i}(\nu^*) \cup \{i\}\right\}$.

Considering a permutation $\hat{\psi}_i \in \Psi_i$ that is obtained from $\tilde{\psi}_i$ by switching a member of $M_{\tilde{\psi}_i}(\nu^*)$, who has all her links within $M_{\tilde{\psi}_i}(\nu^*)$, with $l \in N \setminus \left\{M_{\tilde{\psi}_i}(\nu^*) \cup \{i\}\right\}$, who has a link $lm \in g|_{N\setminus\left\{M_{\tilde{\psi}_i}(\nu^*)\cup\{i\}\right\}}$, contradicts the assumption that $\tilde{\psi}_i$ yielded the minimal number of j with $\eta_j(g|_{M_{\psi_i}(\nu^*)}) = \eta$. Hence, there exists a permutation $\psi_i^* \in \Psi_i$ such that for all $j \in M_{\psi_i}(\nu^*)$ we have $|M_{\psi_i^*}(\nu^*) \cap N_j| \leq |M_{\psi_i^*}(\nu^*) \cap N_i| - 1$, implying that (26) holds.

Finally, choosing $P_i := M_{\psi_i^*}(\nu^*)$ yields first that trivially (12) is satisfied. Moreover, because of minimization we have that (12) cannot be satisfied by any subset of P_i. Thus $\big((|P_i \cap N_i| - 1)\gamma + (|P_i| - 1)\beta\big) < \frac{1}{2}\big(\beta(n-1) + \gamma\eta\big)$ and since (26) holds for $P_i = M_{\psi_i^*}(\nu^*)$, we get that (13) is satisfied. Hence, both conditions of Theorem 1 are satisfied by choosing a punishment set $P_i := M_{\psi_i^*}(\nu^*)$ for every $i \in N$, implying that there exists a WRP equilibrium.

Note that we have shown here a slightly general result since we have shown that the WRP condition holds for all punishers. □

Proof of Lemma 1. Suppose a player j has deviated and players P_j are called upon to punish. When a player i joins the punishment group P_j, the marginal effect on total welfare $\mathcal{W}$ can be calculated to be

$$\begin{aligned}
\Delta\mathcal{W}(P_j, i) &= -\Big(\beta n + \gamma(\eta_i + 1)\Big)\Big(\beta(n-1) + \gamma\eta_i\Big) - \frac{1}{2}\bigg((\beta+\gamma)^2 - \big(\beta n + \gamma(\eta_i+1)\big)^2\bigg) \\
&= -\Big(\beta n + \gamma(\eta_i + 1)\Big)\Big(\beta(n-1) - \frac{1}{2}\beta n + \gamma\eta_i - \frac{1}{2}\gamma(\eta_i+1)\Big) - \frac{1}{2}(\beta+\gamma)^2 \\
&= -\frac{1}{2}\Big(\beta n + \gamma(\eta_i + 1)\Big)\Big(\beta(n-1) + \gamma\eta_i - \beta - \gamma\Big) - \frac{1}{2}(\beta+\gamma)^2 \\
&= -\frac{1}{2}\Big(\beta(n-1) + \gamma\eta_i\Big)^2 + \frac{1}{2}(\beta+\gamma)\left(\beta n + \gamma(\eta_i+1) - \beta(n-1) + \gamma\eta_i\right) - \frac{1}{2}(\beta+\gamma)^2 \\
&= -\frac{1}{2}\Big(\beta(n-1) + \gamma\eta_i\Big)^2 + \frac{1}{2}(\beta+\gamma)\left(\beta n + \gamma(\eta_i+1) - \beta(n-1) + \gamma\eta_i - \beta - \gamma\right) \\
&= -\frac{1}{2}\Big(\beta(n-1) + \gamma\eta_i\Big)^2
\end{aligned}$$

□

Proof of Proposition 5. Suppose a player j has deviated and players P_j are called upon to punish. Let $i \in N_j$ and $l, m \notin N_i \cup \{i\}$ be such that $f_j(\cdot, P_j \cup \{i\}) \leq f_j(\cdot, P_j \cup \{l, m\})$.

That is, we have that

$$\begin{aligned} & \beta\Big(\beta(n-1)+\gamma\eta_l+\beta(n-1)+\gamma\eta_m\Big) \geq (\beta+\gamma)\Big(\beta(n-1)+\gamma\eta_i\Big) \\ \Leftrightarrow\quad & \beta^2\Big(\beta(n-1)+\gamma\eta_l+\beta(n-1)+\gamma\eta_m\Big)^2 \geq (\beta+\gamma)^2\Big(\beta(n-1)+\gamma\eta_i\Big)^2 \\ \Leftrightarrow\quad & \underbrace{\frac{\beta^2}{(\beta+\gamma)^2}}_{=:a}\Bigg(\underbrace{\Big[\beta(n-1)+\gamma\eta_l\Big]^2+\Big[\beta(n-1)+\gamma\eta_m\Big]^2}_{=:\xi_1}+\underbrace{2\Big[\,\big(\beta(n-1)+\gamma\eta_l\big)\,\big(\beta(n-1)+\gamma\eta_m\big)\,\Big]}_{=:\xi_2}\Bigg) \\ & \geq \underbrace{\Big[\beta(n-1)+\gamma\eta_i\Big]^2}_{=:\xi_3}. \end{aligned}$$

Next, note that $\xi_1 \geq \xi_3$ if $\xi_1 \geq a(\xi_1+\xi_2)$, i.e. $\xi_1 - \frac{a\xi_2}{1-a} \geq 0$. This is equivalent to

$$0 \leq \Big[\beta(n-1)+\gamma\eta_l\Big]^2+\Big[\beta(n-1)+\gamma\eta_m\Big]^2-\frac{2\beta^2\Big[(\beta(n-1)+\gamma\eta_l)(\beta(n-1)+\gamma\eta_m)\Big]}{2\beta\gamma+\gamma^2}.$$

Choosing $\gamma\eta_l = \Big(\beta(n-1)+\gamma\eta_m\Big)\frac{\beta^2}{2\beta\gamma+\gamma^2} - \beta(n-1)$ minimizes the right-hand side and thus above is implied by

$$\Leftarrow 0 \leq \Big[\beta(n-1)+\gamma\eta_m\Big]^2\left[1+\left(\frac{\beta^2}{2\beta\gamma+\gamma^2}\right)^2\right]-2\left(\frac{\beta^2}{2\beta\gamma+\gamma^2}\right)^2\Big[\beta(n-1)+\gamma\eta_m\Big]^2,$$

which in turn is equivalent to $\beta^2 \leq 2\beta\gamma+\gamma^2$. For positive values, we get $\beta \leq (1+\sqrt{2})\gamma$, which concludes the proof. □

References

Dilip Abreu. On the theory of infinitely repeated games with discounting. *Econometrica: Journal of the Econometric Society*, pages 383–396, 1988.

Nizar Allouch. On the private provision of public goods on networks. *Journal of Economic Theory*, 157:527 – 552, 2015. ISSN 0022-0531. doi: http://dx.doi.org/10.1016/j.jet.2015.01.007. URL `http://www.sciencedirect.com/science/article/pii/S0022053115000095`.

Geir B. Asheim and Bjart Holtsmark. Renegotiation-proof climate agreements with full participation: Conditions for pareto-efficiency. *Environmental and Resource Economics*, 43(4):519–533, 2009.

Geir B. Asheim, Camilla Bretteville Froyn, Jon Hovi, and Fredric C. Menz. Regional versus global cooperation for climate control. *Journal of Environmental Economics and Management*, 51(1):93–109, 2006.

Scott Barrett. Self-enforcing international environmental agreements. *Oxford Economic Papers*, pages 878–894, 1994.

Scott Barrett. A theory of full international cooperation. *Journal of Theoretical Politics*, 11(4):519–541, 1999.

Marco Battaglini and Bård Harstad. Participation and duration of environmental agreements. Working Paper 18585, National Bureau of Economic Research, December 2012. URL `http://www.nber.org/papers/w18585`.

Hassan Benchekroun and Ngo Van Long. Collaborative environmental management: A review of the literature. *International Game Theory Review*, 14(04), 2012.

Francis Bloch and Unal Zenginobuz. The effect of spillovers on the provision of local public goods. *Review of Economic Design*, 11(3):199–216, 2007.

Johannes Bollen, Bob van der Zwaan, Corjan Brink, and Hans Eerens. Local air pollution and global climate change: A combined cost-benefit analysis. *Resource and Energy Economics*, 31(3):161–181, 2009.

Yann Bramoullé and Rachel Kranton. Public goods in networks. *Journal of Economic Theory*, 135(1):478–494, 2007.

Sergio Currarini, Carmen Marchiori, and Alessandro Tavoni. Network economics and the environment: insights and perspectives. *FEEM Working Paper No. 6.2014*, 2014.

Engelbert J Dockner and Kazuo Nishimura. Transboundary pollution in a dynamic game model. *Japanese Economic Review*, 50(4):443–456, 1999.

Matthew Elliott and Benjamin Golub. A network approach to public goods. In *Proceedings of the fourteenth ACM conference on Electronic commerce*, pages 377–378. ACM, 2013.

Joseph Farrell and Eric Maskin. Renegotiation in repeated games. *Games and economic behavior*, 1(4):327–360, 1989.

Michael Finus and Bianca Rundshagen. Endogenous coalition formation in global pollution control. *FEEM Working Paper No. 43.2001*, 2001.

Camilla Bretteville Froyn and Jon Hovi. A climate agreement with full participation. *Economics Letters*, 99(2):317–319, 2008.

Drew Fudenberg and Eric Maskin. The folk theorem in repeated games with discounting or with incomplete information. *Econometrica*, pages 533–554, 1986.

Rögnvaldur Hannesson. The coalition of the willing: Effect of country diversity in an environmental treaty game. *The Review of International Organizations*, 5(4):461–474, 2010.

Steffen Jørgensen, Guiomar Martín-Herrán, and Georges Zaccour. Dynamic games in the economics and management of pollution. *Environmental Modeling and Assessment*, 15 (6):433–467, 2010.

Thomas Kühn, Antti-Ilari Partanen, Svante V Henriksson, Tommi Bergman, Anton Laakso, Harri Kokkola, Sami Romakkaniemi, and Ari Laaksonen. Impact on aerosol emissions in china and india on local and global climate. In *EGU General Assembly Conference Abstracts*, volume 15, page 10188, 2013.

Matthew McGinty. International environmental agreements among asymmetric nations. *Oxford Economic Papers*, 59(1):45–62, 2007.

Committee on the Significance of International Transport of Air Pollutants; National Research Council. *Global Sources of Local Pollution: An Assessment of Long-Range Transport of Key Air Pollutants to and from the United States.* The National Academies Press, 2009.

NAD Richards, SR Arnold, MP Chipperfield, A Rap, SA Monks, MJ Hollaway, G Miles, and R Siddans. The mediterranean summertime ozone maximum: Global emission sensitivities and radiative impacts. *Atmospheric Chemistry and Physics*, 13(5):2331–2345, 2013.

Santiago J Rubio and Alistair Ulph. Self-enforcing international environmental agreements revisited. *Oxford economic papers*, 58(2):233–263, 2006.

Zili Yang. Negatively correlated local and global stock externalities: tax or subsidy? *Environment and Development Economics*, 11(03):301–316, 2006.

Sang-Seung Yi and Hyukseung Shin. Endogenous formation of research coalitions with spillovers. *International Journal of Industrial Organization*, 18(2):229–256, 2000.