

Eichner, Thomas; Schopf, Mark

Article — Published Version

Self-Enforcing Environmental Agreements with the Optimal Intensity of Cooperation

Environmental and Resource Economics

Provided in Cooperation with:

Springer Nature

Suggested Citation: Eichner, Thomas; Schopf, Mark (2025) : Self-Enforcing Environmental Agreements with the Optimal Intensity of Cooperation, Environmental and Resource Economics, ISSN 1573-1502, Springer Netherlands, Dordrecht, Vol. 88, Iss. 2, pp. 249-278, <https://doi.org/10.1007/s10640-024-00927-1>

This Version is available at:

<https://hdl.handle.net/10419/318624>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.


If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<http://creativecommons.org/licenses/by/4.0/>



Self-Enforcing Environmental Agreements with the Optimal Intensity of Cooperation

Thomas Eichner¹ · Mark Schopf¹ 

Accepted: 17 September 2024 / Published online: 9 January 2025
© The Author(s) 2025

Abstract

This paper analyzes international environmental agreements in three-stage games consisting of a membership subgame, the signatories' decision on the intensity of cooperation, and an emissions subgame. Signatories may act as Stackelberg leaders or play Nash. In the Stackelberg game, the highest intensity of cooperation between signatories is optimal. In the Nash game, a moderate or the highest intensity of cooperation is optimal if emissions are strategic substitutes. In this case the equilibrium emissions and the stable coalition of the Nash game with optimal cooperation intensity are identical to the equilibrium emissions and the stable coalition of the Stackelberg game with the highest cooperation intensity. Finally, we apply our results to coalition formation games of the literature. In the applied Nash games, optimizing with respect to the cooperation intensity may enlarge the stable coalition up to the grand coalition which implements the first best.

Keywords Optimal intensity of cooperation · Stable coalition · Nash · Stackelberg

JEL Classification C71 · F55 · Q54

1 Introduction

Effective international environmental agreements (IEAs) with many countries are necessary to keep global warming well below two degrees Celsius and to protect the community against catastrophic damages. There is a literature dating back to Hoel (1992), Carraro and Siniscalco (1993) and Barrett (1994) that analyzes IEAs by means of two-stage membership games developed in cartel theory (d'Aspremont et al. 1983). In the first stage of these games the coalition is formed and in the second stage countries decide on emissions and mitigation, respectively. A subgame perfect equilibrium is reached when the coalition is stable, i.e.

✉ Mark Schopf
mark.schopf@fernuni-hagen.de

Thomas Eichner
thomas.eichner@fernuni-hagen.de

¹ Department of Economics, University of Hagen, Universitätsstr. 41, Hagen 58097, Germany

when there is no incentive to leave or to join the coalition. In this literature it is assumed that signatories maximize the sum of their welfares when choosing mitigation, and that either a Stackelberg game or a Nash game is played between members and non-members of the coalition.

The most prominent IEAs are the Kyoto Protocol and the Paris Agreement. In contrast to the “top-down” approach of the Kyoto Protocol, the Paris Agreement is a “bottom-up” approach which requires each country to make voluntary pledges (nationally determined contributions) towards achieving the long-term temperature goal of two degrees Celsius. Harstad (2023a, 2023b) and Colombo et al. (2022) argue that the cooperation intensity of signatories is low or moderate in the Paris Agreement, whereas it is high in the Kyoto Protocol. The coefficient of cooperation measures the cooperation intensity and ranges from zero to one. At the lowest cooperation intensity, a signatory is purely self-interested and maximizes its own welfare. At the highest cooperation intensity, signatories maximize the sum of their welfares.

Dimitrov et al. (2019, p. 6) describe the Paris Agreement as “shallow coordination” and the initial pledges as “not negligible yet not nearly ambitious enough”. The initial pledges are insufficient to limit global warming below two degree Celsius (cf. Rogelj et al. (2016)). Pledges are renegotiated every five years and scholars hope future pledges will be more ambitious (Falkner 2016). UNFCCC (2015), which rules the Paris Agreement, requires the signatories to enhance international cooperation for climate action in general, and in particular to strengthen cooperation on mitigation, on adaption and on technology development and transfer. Thus it is anchored in the Paris Agreement that countries should intensify cooperation over time. From the perspective of a signatory increasing the intensity of cooperation comes at the benefit of reducing global emissions and at the cost of reducing own emissions or tantamount at a consumption loss.

This trade-off may induce signatories to search for and choose the optimal cooperation intensity. Although the cooperation intensity cannot be contractually specified in the IEA and the cooperation intensity is not a usual decision variable, we assume that signatories behave as if they adjust and choose their cooperation intensity. In order to implement that as-if behavior, we endogenize the cooperation intensity and let the signatories maximize with respect to the intensity. Determining the best cooperation intensity answers the question what the Paris Agreement may achieve in the future when signatories optimize their cooperation intensity.

To the best of our knowledge, the intensity of cooperation has only been applied by Harstad (2023a, 2023b) and Colombo et al. (2022) in Nash games. Harstad¹ (2023b) analyzes a dynamic bargaining game denoted as pledge-and-review bargaining game in which countries decide whether to participate in an IEA, signatories decide on the duration of the IEA, and countries choose emissions and investments in renewables. Harstad (2023b) argues that features of the Paris agreement can be explained by his pledge-and-review bargaining with low cooperation intensity and features of the Kyoto Protocol by the pledge-and-review bargaining with high cooperation intensity. The cooperation intensity in the pledge-and-review bargaining is determined by the countries’ uncertainty concerning the discount rate and cannot be influenced by the coalition countries themselves. The stable coalition is the larger and global welfare is the higher the weaker is the cooperation intensity

¹ The pledge-and-review bargaining game used in Harstad (2023b) is microfounded as a non-cooperative game in Harstad (2023a).

which explains the participation of many countries in the Paris Agreement and the participation of few countries in the Kyoto Protocol. In contrast to Harstad (2023a, 2023b), we consider simple static games in which the cooperation intensity is the choice variable of the signatories. Colombo et al. (2022) study a dynamic two-stage game with a stock of a global public good and with exogenous cooperation intensity. In the first stage, countries decide whether to participate in an IEA. In the second stage, countries choose their contribution to a public good. Colombo et al. (2022) show that the coalition size may be large and increases over time, and that the stable coalition and discounted global welfare are the larger the lower the cooperation intensity.

Our paper is also related to a literature in which coalition countries do not fully internalize the climate externalities within the coalition. Finus and Rundshagen (1998) and Altamirano-Cabrera et al. (2008) apply the smallest common denominator bargaining, which was introduced by Endres (1997) to asymmetric countries that agree on the lowest quota (or tax) proposal, and employ that the smallest common denominator bargaining can stabilize the grand coalition. In a repeated game, Barrett (2002) shows that a consensus treaty with all countries (grand coalition) can be sustained, if the countries' emissions reductions are modest. Finus and Maus (2008) endogenize the degree of modesty and analyze consensus treaties of all possible coalitions in a two-stage membership game with an additional pre-stage. They find that modesty may lead to higher participation and more successful treaties.

The present paper analyzes the intensity of cooperation in a general static setting. In contrast to Harstad (2023a, 2023b) and Colombo et al. (2022), the cooperation intensity is determined by the signatories via maximizing their welfare. The results of our games depend on whether a Nash or Stackelberg game is played and whether emissions are strategic substitutes or strategic complements. In the Nash game, signatories account for the emissions leakage caused by non-signatories when choosing the cooperation intensity. If emissions are strategic substitutes [complements], the optimal cooperation intensity is weakly lower than [equal to] the highest cooperation intensity (Proposition 1). By contrast, in the Stackelberg game signatories choose the highest cooperation intensity (Proposition 2). It turns out that the equilibrium emissions and the stable coalition of the Nash game with optimal cooperation intensity coincide with the equilibrium emissions and the stable coalition of the Stackelberg game with highest cooperation intensity if and only if emissions are strategic substitutes (Proposition 3).

Our Propositions 1–3 are applied to a variety of games analyzed in the literature under the assumption of the exogenously given highest cooperation intensity among signatories. In emissions and abatement games, an emissions and trade game, an adaptation and mitigation game and a fishing game strategies are strategic substitutes. In these games with strategic substitutes, optimizing with respect to the cooperation intensity in Nash games may enlarge the size of the stable coalition up to the grand coalition whereby the latter implements the first best. In a mitigation and adaptation game in which strategies are strategic complements optimizing yields the highest cooperation level both in Nash and Stackelberg games and optimizing does not change the equilibrium emissions and the stable coalition compared to the literature which a priori assumed that signatories cooperate with the highest intensity of cooperation.

The present paper is organized as follows: In Sect. 2 we introduce the welfare function and the intensity of cooperation. Subsection 3.1 analyzes optimal cooperation intensity in the Nash game and Subsect. 3.2 optimal cooperation intensity in the Stackelberg game.

Subsection 3.3 compares the Nash and Stackelberg equilibria. Section 4 applies our propositions to existing games and results of the literature. Section 5 concludes.

2 The Welfare Function and the Intensity of Cooperation

Consider an economy with $n \geq 3$ identical countries. The welfare of country $i \in N = \{1, \dots, n\}$ is represented by the at least three times differentiable function $W : \mathbb{R}_+ \times \mathbb{R}_+ \rightarrow \mathbb{R}$ satisfying

$$w_i = W(e_i, E), \quad (1)$$

where $e_i \in \mathbb{R}_+$ are country i 's emissions and $E = \sum_{j \in N} e_j$ are aggregate or global emissions. Country i benefits from own emissions and these emissions impose a negative externality on country $j \neq i$, which is formally reflected by the partial derivatives $W_e > 0$ and $W_E < 0$.²

As motivated in the introduction, we follow Colombo et al. (2022) and Harstad (2023a, 2023b) and assume that each signatory's choice of emissions is partly group-oriented and partly self-interested. Denoting the set of signatories by $C := \{1, \dots, m\}$, signatory i 's objective function is

$$W(e_i, E) + \gamma \sum_{j \in C \setminus i} W(e_j, E), \quad (2)$$

where $\gamma \in [0, 1]$ reflects the cooperation intensity and is referred to as coefficient of cooperation. In the polar case $\gamma = 1$, signatories are purely group-oriented and maximize the sum of the signatories' welfares. In the polar case $\gamma = 0$, signatories are purely self-interested and maximize their own welfare.

3 The Games

In this section, we consider the following three-stage games: At stage 1, countries play a participation subgame and decide whether to join a climate coalition or to stay outside the coalition. At stage 2, signatories jointly choose the coefficient of cooperation γ and at stage 3, all countries play an emissions subgame and choose their emissions. The games are solved by backward induction. We distinguish between two kinds of emissions subgames. At the Nash game signatories and non-signatories set their emissions simultaneously, whereas at the Stackelberg game signatories set emissions prior to non-signatories.

3.1 Nash Game

We begin with the Nash game. At the emissions subgame of stage 3 both non-signatories and signatories play Nash. Each non-signatory $i \in F = N \setminus C$ maximizes its welfare func-

²The results of section 3 neither rest on the interpretation of emissions nor on the signs of the partial derivatives.

tion (1) with respect to e_i for given emissions of the other non-signatories and for given emissions of signatories which yields the first-order condition

$$W_e(e_i, E) + W_E(e_i, E) = 0 \quad \forall i \in F \quad (3)$$

and the second-order condition³

$$W_{ee}^f + W_{Ee}^f + W_{eE}^f + W_{EE}^f < 0. \quad (4)$$

The non-signatories' first-order condition (3) requires the marginal benefits of emissions, represented by $W_e > 0$, and the marginal costs of emissions, formally reflected by $-W_E > 0$, to be equal. Setting $e_i = e_f$ for all $i \in F$, accounting for $E = (n - m)e_f + \sum_{j \in C} e_j$, and solving (3) with respect to $(n - m)e_f$ we obtain the non-signatories' aggregate best-response function $(n - m)e_f = \mathcal{RF} \left(\sum_{j \in C} e_j \right)$ with⁴

$$\mathcal{RF}' = - \frac{(n - m) \left(W_{ee}^f + W_{EE}^f \right)}{W_{ee}^f + W_{Ee}^f + (n - m) \left(W_{eE}^f + W_{EE}^f \right)}. \quad (5)$$

Each signatory $i \in C$ maximizes the objective function (2) subject to $E = \sum_{j \in F} e_j + e_i + \sum_{j \in C \setminus i} e_j$ with respect to e_i . The first-order condition is given by

$$W_e(e_i, E) + [1 + \gamma(m - 1)] W_E(e_i, E) = 0 \quad \forall i \in C \quad (6)$$

and the second-order condition reads⁵

$$W_{ee}^c + [1 + \gamma(m - 1)] W_{Ee}^c + W_{eE}^c + [1 + \gamma(m - 1)] W_{EE}^c < 0. \quad (7)$$

The signatories' first-order condition (6) also balances the marginal benefits and the marginal costs of emissions. The marginal benefits are equal to those of the non-signatories but the marginal costs are different. Whereas a non-signatory accounts only for the impact of its emissions on its own marginal costs, the signatories internalize the share of externalities $\gamma(m - 1)$ imposed on other signatories. Setting $e_i = e_c$ for all $i \in C$, accounting for $E = me_c + \sum_{j \in F} e_j$, and solving (6) yields the signatories' aggregate best-response function $me_c = \mathcal{RC} \left(\sum_{j \in F} e_j \right)$ with⁶

$$\mathcal{RC}' = - \frac{m [W_{ee}^c + (1 + \gamma(m - 1)) W_{EE}^c]}{W_{ee}^c + (1 + \gamma(m - 1)) W_{Ee}^c + m [W_{eE}^c + (1 + \gamma(m - 1)) W_{EE}^c]}. \quad (8)$$

³We use the notation $W_{ee}^f := \frac{\partial^2 W(e_i, E)}{(\partial e_i)^2}$, $W_{Ee}^f := \frac{\partial^2 W(e_i, E)}{(\partial E) \partial e_i}$, $W_{eE}^f := \frac{\partial^2 W(e_i, E)}{\partial e_i \partial E}$ for $i \in F$ to avoid clutter.

⁴Total differentiation of (3) with $E = (n - m)e_f + \sum_{j \in C} e_j$ and rearranging terms yields (5).

⁵We use the notation $W_{ee}^c := \frac{\partial^2 W(e_i, E)}{(\partial e_i)^2}$, $W_{Ee}^c := \frac{\partial^2 W(e_i, E)}{(\partial E) \partial e_i}$, $W_{eE}^c := \frac{\partial^2 W(e_i, E)}{\partial e_i \partial E}$ for $i \in C$ to avoid clutter.

⁶Total differentiation of (6) with $E = me_c + \sum_{j \in F} e_j$ and rearranging terms yields (8).

The first-order conditions (3) and (6) jointly determine the Nash equilibrium of the emissions subgame in dependence of the cooperation coefficient γ and of the coalition size m , formally

$$\begin{aligned} e_i &= e_c = \hat{e}^c(\gamma, m) \quad \forall i \in C, & e_i &= e_f = \hat{e}^f(\gamma, m) \quad \forall i \in F, \\ E &= \hat{E}(\gamma, m) = m\hat{e}^c(\gamma, m) + (n - m)\hat{e}^f(\gamma, m), \end{aligned} \quad (9)$$

where the hat marks Nash equilibrium values of the emissions subgame.

In the following, we need several assumptions. Our first assumption requires⁷

Assumption 1 The welfare function $W(e_i, E)$ fulfills the properties $W_{ee} < 0$, $W_{eE} \leq 0$, $W_{ee} + 2nW_{eE} + n^2W_{EE} < 0$.

Assumption 1 ensures that the second-order conditions (4) and (7) are satisfied for all $m \in \{2, \dots, n\}$ and $\gamma \in [0, 1]$ in the emissions subgame. In Lemma 1 of the Appendix, we show that $W_{ee} + 2nW_{eE} + n^2W_{EE} < 0$ is necessary for the existence and uniqueness of the Nash equilibrium in the emissions subgame for $m = n$ and $\gamma = 1$. Moreover, Assumption 1 is sufficient for the existence and uniqueness of the Nash equilibrium in the emissions subgame for all $m \in \{2, \dots, n\}$ and $\gamma \in [0, 1]$.

Closer inspection of (5) and (8) shows that Assumption 1 also implies that the slopes of the aggregate best-response functions satisfy $\mathcal{R}\mathcal{F}' > -1$ and $\mathcal{R}\mathcal{C}' > -1$, and that the sign of $W_{eE} + W_{EE}$ determines whether non-signatories' emissions are strategic complements or substitutes. Emissions are strategic substitutes if $W_{eE} + W_{EE} < 0$, and complements if $W_{eE} + W_{EE} > 0$. In the former case, the best response curve is downward-sloping ($\mathcal{R}\mathcal{F}' \in [-1, 0]$), in the latter case, it is upward-sloping ($\mathcal{R}\mathcal{F}' > 0$).

At stage 2, the coalition countries maximize their welfare sum $mW(e_c, E)$ subject to (9), $\hat{e}^f(\gamma, m) = \mathcal{R}\mathcal{F}[m\hat{e}^c(\gamma, m)]$ and $\gamma \in [0, 1]$ with respect to γ .⁸ Solving the associated Lagrangean $\mathcal{L} = mW[\hat{e}^c(\gamma, m), \hat{E}(\gamma, m)]$ yields the first-order condition⁹

$$\mathcal{L}_\gamma = m\hat{e}_\gamma^c \cdot W_e(e_i, E) + m\hat{E}_\gamma \cdot W_E(e_i, E) \geq 0, \quad \gamma(1 - \gamma)\mathcal{L}_\gamma = 0 \quad \forall i \in C. \quad (10)$$

In Lemma 2 of the Appendix, we show that Assumption 1 implies $\hat{e}_\gamma^c \cdot W_e < 0$ and $\hat{E}_\gamma \cdot W_E > 0$. Due to $W_e > 0$ and $W_E < 0$, a higher cooperation intensity reduces both the signatories' emissions and the global emissions, formally $\hat{e}_\gamma^c < 0$ and $\hat{E}_\gamma < 0$. In (10), the term $m\hat{e}_\gamma^c \cdot W_e$ reflects the marginal costs and the term $m\hat{E}_\gamma \cdot W_E > 0$ captures the marginal benefits of the cooperation intensity. In an interior solution, these marginal benefits and costs are balanced. Accounting for¹⁰ $\hat{E}_\gamma = m(1 + \mathcal{R}\mathcal{F}')\hat{e}_\gamma^c$ we obtain from (10)

$$W_e(e_i, E) + m(1 + \mathcal{R}\mathcal{F}')W_E(e_i, E) \geq 0, \quad \gamma(1 - \gamma)\mathcal{L}_\gamma = 0 \quad \forall i \in C. \quad (11)$$

⁷Our results also hold if $W_{ee} < 0$, $W_{eE} \leq 0$ is replaced by $W_{ee} \leq 0$, $W_{eE} < 0$.

⁸Since the non-signatories choose their emissions at stage 3, the signatories take their aggregate best response into account when they choose the intensity of cooperation at stage 2.

⁹We use the notation $\hat{e}_\gamma^c := \frac{\partial \hat{e}^c(\gamma, m)}{\partial \gamma}$, $\hat{E}_\gamma := \frac{\partial \hat{E}(\gamma, m)}{\partial \gamma}$ to avoid clutter.

¹⁰Observe that $\hat{E}_\gamma = n\hat{e}_\gamma^c$ if $m = n$. The formal treatment of $m = n$ can be included in the subsequent analysis by assuming $\mathcal{R}\mathcal{F}' = 0$ if $m = n$.

The second-order condition for an interior solution reads

$$W_{ee}^c + m(1 + \mathcal{R}\mathcal{F}') W_{eE}^c + m(1 + \mathcal{R}\mathcal{F}') [W_{Ee}^c + m(1 + \mathcal{R}\mathcal{F}') W_{EE}^c] + m^2 \mathcal{R}\mathcal{F}'' \cdot W_E^c < 0. \quad (12)$$

Combining (6) and (11) and solving for γ establishes

$$\gamma = G(m) := \begin{cases} 1 & \text{if } \Gamma(m) \geq 1, \\ \Gamma(m) & \text{if } \Gamma(m) \in]0, 1[, \\ 0 & \text{if } \Gamma(m) \leq 0, \end{cases} \quad (13)$$

where $\Gamma(m) := 1 + \frac{m}{m-1} \mathcal{R}\mathcal{F}'$. The grand coalition chooses $\gamma = 1$ due to $\Gamma(n) = 1$. A coalition of size $m \in \{2, \dots, n-1\}$ chooses the corner solution $\gamma = 1$ if $\Gamma(m) \geq 1$, the interior solution $\gamma \in]0, 1[$ if $\Gamma(m) \in]0, 1[$, and the corner solution $\gamma = 0$ if $\Gamma(m) \leq 0$. If emissions are strategic complements for non-signatories, then $\Gamma(m) \geq 1$ and the coalition chooses the corner solution $\gamma = 1$. If the coefficient of cooperation γ were unconstrained, in case of strategic complements the coalition would choose $\gamma > 1$ to benefit from negative emissions leakage. Leakage is negative if non-signatories decrease their emissions as response to an emissions reduction of the signatories ($\mathcal{R}\mathcal{F}' > 0$). By contrast, if emissions are strategic substitutes for non-signatories, emissions leakage is positive ($\mathcal{R}\mathcal{F}' \in]-1, 0[$) and the coalition chooses $\gamma < 1$. Since smaller coalitions are ceteris paribus accompanied by a larger emissions leakage ($\partial \mathcal{R}\mathcal{F}' / \partial m > 0$) and a lower cooperation intensity ($\partial \Gamma(m) / \partial m > 0$), the corner solution $\gamma = 0$ could be optimal for small coalitions. In case of the grand coalition ($m = n$) the optimal cooperation intensity is $\gamma = 1$. Making use of $G(m)$ yields the welfare functions

$$\begin{aligned} \hat{W}^c(m) &= W[\hat{e}^c(G(m), m), \hat{E}(G(m), m)], \\ \hat{W}^f(m) &= W[\hat{e}^f(G(m), m), \hat{E}(G(m), m)]. \end{aligned} \quad (14)$$

Our next assumption requires

Assumption 2 The welfare function $W(e_i, E)$ fulfills the property $\mathcal{R}\mathcal{F}'' \cdot W_E^c \leq 0$.

In Lemma 3 of the Appendix, we show that Assumptions 1 and 2 are sufficient for the second-order condition (12) to be satisfied for any $m \in \{2, \dots, n\}$.

At stage 1, we turn to the participation game. Countries decide whether to participate or not in an IEA. A coalition of size m^* is stable if it satisfies the internal stability condition $\hat{W}^c(m^*) \geq \hat{W}^f(m^* - 1)$ and the external stability condition $\hat{W}^c(m^* + 1) < \hat{W}^f(m^*)$ of d'Aspremont et al. (1983). The subgame perfect equilibrium (SPE) of the Nash game then is characterized by the stable coalition m^* , the coefficient of cooperation $\gamma^* = G(m^*) \in [0, 1]$, equilibrium emissions $e_c^* = \hat{e}^c(\gamma^*, m^*)$ and $e_f^* = \hat{e}^f(\gamma^*, m^*)$, and equilibrium welfares $w_c^* = \hat{W}^c(m^*)$ and $w_f^* = \hat{W}^f(m^*)$.

In Lemma 4 of the Appendix we show that

$$\hat{W}^f(m) > \hat{W}^c(m) > w^{\text{BAU}} \iff G(m) > 0, \quad (15)$$

$$\hat{W}^c(m) = w^{\text{BAU}} = \hat{W}^f(m) \iff G(m) = 0, \quad (16)$$

where w^{BAU} is the countries' welfare level in the business as usual (BAU).¹¹ Due to $\hat{W}^c(m+1) \geq w^{\text{BAU}} = \hat{W}^f(m)$, the external stability condition is violated for $G(m) = 0$ and there does not exist a SPE of the Nash game. Consequently, $G(m^*) > 0$. In particular, we prove in the Appendix

Proposition 1 Under Assumptions 1 and 2, the SPE of the Nash game with optimal cooperation intensity is characterized by

- (i) (a) $\gamma^* \in]0, 1[$ and $2 \leq m^* < n$, if emissions are strategic substitutes or
- (b) $\gamma^* = 1$ and $m^* = n$, if emissions are strategic substitutes;
- (ii) $\gamma^* = 1$ and $m^* \geq 2$, if emissions are strategic complements.

If emissions are strategic substitutes (see Proposition 1 (i)), either the signatories' best choice is the highest cooperation intensity and the grand coalition is stable, or the (positive) emissions leakage causes signatories to choose the moderate cooperation intensity $\gamma^* < 1$ and some smaller coalition is stable. In case of $\Gamma(m) < 1$, for any given coalition size m the signatories' [non-signatories'] welfare is higher [lower] in the Nash game with optimal cooperation intensity than in a Nash game in which signatories take the highest cooperation coefficient $\gamma = 1$ as exogenously given. This relaxes the internal stability condition and tightens the external stability condition, such that the equilibrium coalition is (weakly) larger in the Nash game with optimal cooperation intensity than in the Nash game with exogenous cooperation intensity $\gamma = 1$. If emissions are strategic complements (see Proposition 1 (ii)), the negative emissions leakage induces signatories to choose the corner solution $\gamma^* = 1$. In that case the equilibrium emissions and the stable coalition of the Nash game with optimal cooperation intensity $\gamma^* = 1$ and the equilibrium emissions and the stable coalition of the Nash game with exogenous cooperation intensity $\gamma = 1$ coincide.

3.2 Stackelberg Game

In this subsection, we turn to the Stackelberg game. Signatories act as Stackelberg leaders and non-signatories as Stackelberg followers in the emissions subgame of stage 3. First, the signatories as Stackelberg leaders and then the non-signatories as Stackelberg followers choose emissions. Applying backward induction to stage 3 the non-signatories' emissions are still determined by the first-order condition (3) which yields the non-signatories' aggregate best-response function $(n-m)e_f = \mathcal{RF}\left(\sum_{j \in C} e_j\right)$, whereas the signatories now decide on emissions taking the non-signatories' aggregate best response into account. More precisely, for given $\gamma \in [0, 1]$ signatory $i \in C$ maximizes the objective function (2) subject

¹¹Formally, BAU ensues when no coalition exists and all countries play Nash against each other, i.e. $\gamma = 1$ and $m = 1$. The associated BAU emissions are $e^{\text{BAU}} = \hat{e}^f(1, 1) = \hat{e}^c(1, 1)$ and BAU welfare levels are $w^{\text{BAU}} = W(e^{\text{BAU}}, ne^{\text{BAU}})$.

to $E = \mathcal{RF} \left(\sum_{j \in C} e_j \right) + e_i + \sum_{j \in C \setminus i} e_j$ with respect to e_i . The first-order condition is given by

$$W_e(e_i, E) + [1 + \gamma(m-1)] (1 + \mathcal{RF}') W_E(e_i, E) = 0 \quad \forall i \in C \quad (17)$$

and the second-order condition reads

$$W_{ee}^c + (1 + \mathcal{RF}') W_{eE}^c + [1 + \gamma(m-1)] (1 + \mathcal{RF}') [W_{EE}^c + (1 + \mathcal{RF}') W_{EE}^c] + [1 + \gamma(m-1)] \mathcal{RF}'' \cdot W_E^c < 0. \quad (18)$$

Recall that $\mathcal{RF}' \in [-1, 0]$ if emissions are strategic substitutes, and $\mathcal{RF}' > 0$ if emissions are strategic complements. The signatories' first-order condition (17) requires the marginal benefits of emissions, represented by $W_e > 0$, and the marginal costs of emissions, formally reflected by $-[1 + \gamma(m-1)] (1 + \mathcal{RF}') W_E > 0$, to be equal. When setting e_i , as in the Nash game signatory i internalizes the share $\gamma(m-1)$ of negative externalities levied on fellow signatories which ceteris paribus reduces signatories' emissions relative to non-signatories' emissions. In addition, the signatories as Stackelberg leaders also account for the response of non-signatories which gives the signatories a first-mover advantage and ceteris paribus leads to an increase [decrease] of signatories' emissions relative to non-signatories' emissions in case of strategic substitutes [complements]. The first-order conditions (3) and (17) jointly determine the Stackelberg equilibrium of the emissions subgame in dependence of the cooperation coefficient γ and the coalition size m , formally

$$\begin{aligned} e_i &= e_c = \tilde{e}^c(\gamma, m) \quad \forall i \in C, & e_i &= e_f = \tilde{e}^f(\gamma, m) \quad \forall i \in F, \\ E &= \tilde{E}(\gamma, m) = m\tilde{e}^c(\gamma, m) + (n-m)\tilde{e}^f(\gamma, m), \end{aligned} \quad (19)$$

where the tilde marks Stackelberg equilibrium values in the emissions subgame.

In Lemma 5 of the Appendix, we show that Assumptions 1 and 2 are sufficient both for the second-order condition (18) to be satisfied for all $m \in \{2, \dots, n\}$ and $\gamma \in [0, 1]$, and for the existence and uniqueness of the Stackelberg equilibrium in the emissions subgame.

At stage 2, coalition countries choose the cooperation coefficient γ . When doing so, they maximize the sum of the coalition countries' welfare, formally $mW(e_c, E)$, with respect to γ , subject to (19), $\tilde{e}^f(\gamma, m) = \mathcal{RF} [m\tilde{e}^c(\gamma, m)]$ and $\gamma \in [0, 1]$.¹² The first-order condition yields

$$m\tilde{e}_\gamma^c \cdot W_e(e_i, E) + m\tilde{E}_\gamma \cdot W_E(e_i, E) = 0 \quad \forall i \in C. \quad (20)$$

Taking advantage of $\tilde{E}_\gamma = m(1 + \mathcal{RF}') \tilde{e}_\gamma^c$ turns the first-order condition (20) into

$$W_e(e_i, s) + m(1 + \mathcal{RF}') \cdot W_E(e_i, E) = 0 \quad \forall i \in C, \quad (21)$$

¹² Since the non-signatories choose their emissions at stage 3, the signatories take their aggregate best response into account when they choose the intensity of cooperation at stage 2.

where the second-order condition is given by (12) and is satisfied due to Assumptions 1 and 2. Comparing (21) and (17) we infer $\gamma = 1$. Making use of that information yields the welfare functions

$$\tilde{W}^c(m) = W[\tilde{e}^c(1, m), \tilde{E}(1, m)], \quad \tilde{W}^f(m) = W[\tilde{e}^f(1, m), \tilde{E}(1, m)]. \quad (22)$$

At stage 1, applying the internal and external stability condition to $\tilde{W}^c(m)$ and $\tilde{W}^f(m)$ determines the stable coalition m^* , the cooperation coefficient $\gamma^* = 1$, equilibrium emissions $e_c^* = \tilde{e}^c(1, m^*)$ and $e_f^* = \tilde{e}^f(1, m^*)$, and equilibrium welfares $w_c^* = \tilde{W}^c(m^*)$ and $w_f^* = \tilde{W}^f(m^*)$.¹³ We summarize the results in¹⁴

Proposition 2 Under Assumptions 1 and 2, the SPE of the Stackelberg game with optimal cooperation intensity is characterized by $\gamma^* = 1$ and $m^* \geq 2$.

The preceding analysis shows that the signatories' best choice is the cooperation coefficient $\gamma^* = 1$. Hence, countries cannot improve in comparison to the standard Stackelberg games of the literature in which the signatories take the cooperation coefficient $\gamma = 1$ as exogenously given.

3.3 Comparison of Nash and Stackelberg Games

Finally, we compare the SPE of the Nash game with the SPE of the Stackelberg game. In case of strategic substitutes, $\gamma^* \in]0, 1]$ is an interior solution in the Nash game (Proposition 1(i)), such that the first-order condition (11) holds with equality. Since (11) with equality is identical to the first-order condition (21) in the Stackelberg game, the equilibrium emissions and the stable coalition of the Nash game coincide with the equilibrium emissions and the stable coalition of the Stackelberg game. In case of strategic complements, the first-order conditions (11) and (21) are different, such that the SPE of the Nash game does not coincide with the SPE of the Stackelberg game. In the Appendix we prove

Proposition 3 Suppose that $m^* < n$.¹⁵ Under Assumptions 1 and 2, the equilibrium emissions and the stable coalition of the Nash game with the optimal intensity of cooperation are identical to the equilibrium emissions and the stable coalition of the Stackelberg game with the highest intensity of cooperation if and only if emissions are strategic substitutes.

In case of strategic substitutes, choosing the cooperation coefficient γ enables a coalition playing Nash against non-signatories to take advantage like a Stackelberg leader. In the Stackelberg game that advantage consists of a first-mover advantage whereas in the Nash game that advantage consists of choosing a moderate cooperation intensity. According to Proposition 3 these strategic advantages are of equal size such that the equilibrium emis-

¹³ Without risk of confusion the SPE of the Stackelberg game is also marked by an asterisk.

¹⁴ Note that \tilde{e}^c maximizes $\tilde{W}^c(m)$ for a given coalition size, such that $\tilde{W}^c(m) \geq w^{\text{BAU}}$ holds for all $m = \{2, \dots, n\}$. Consequently, $m = 2$ is internally stable, which implies $m^* \geq 2$.

¹⁵ Suppose that $m^* = n$. Then the SPE of the Nash game and the SPE of the Stackelberg game coincide. In both games signatories choose $\gamma^* = 1$ and the stable grand coalition implements the socially optimal emissions. This holds regardless of whether emissions are strategic substitutes or strategic complements.

sions and the stable coalition are identical in the Nash game with optimal cooperation intensity and in the Stackelberg game with the highest cooperation intensity among signatories in case of strategic substitutes.

In case of strategic complements, a coalition playing Nash against non-signatories would prefer a cooperation coefficient greater than one, but chooses $\gamma^* = 1$ because $\gamma > 1$ is not feasible. For $\gamma^* = 1$ and $m^* < n$ the equilibrium emissions of the Nash game with optimal cooperation intensity are different from the equilibrium emissions of the Stackelberg game with the highest cooperation intensity. In particular, the share of externalities which the signatories internalize is smaller in the Nash game than in the Stackelberg game for any given coalition. The associated internalization shares are m^* in the Nash game and $m^*(1 + \mathcal{RF}') > m^*$ in the Stackelberg game. Each non-signatory's welfare increases with the internalization share. Hence, the condition for the grand coalition to be stable is weaker in the Nash game than in the Stackelberg game, formally $\hat{W}^c(n) - \hat{W}^f(n-1) > \tilde{W}^c(n) - \tilde{W}^f(n-1)$.

4 Applications

In this section, we briefly report on models and the associated results to which Propositions 1-3 can be applied. This literature restricts attention to Nash and Stackelberg games with exogenous cooperation intensity $\gamma = 1$ among signatories. Propositions 1-3 are applicable if Assumptions 1 and 2 are satisfied. Whereas Assumption 1 can be easily checked for concrete welfare functions, Assumption 2 does not relate to the primitives imposed on welfare functions. To have a more tractable assumption, we prove in Lemma 7 of the Appendix that Assumption 2 can be replaced by

Assumption 3 The welfare function $W(e_i, E)$ fulfills the properties $W_{eeE} = W_{eEE} = 0$, $W_E W_{eee} \leq 0$ and $W_E W_{EEE} \leq 0$.

In all games presented below except in the adaptation-mitigation game of subsection 4.3 strategies are strategic substitutes, such that the equilibrium emissions and the stable coalition of the Nash game with optimal cooperation intensity are identical to the equilibrium emissions and the stable coalition of the Stackelberg game according to Proposition 3.¹⁶

4.1 Emissions and Abatement

We begin with the most frequently used emissions game in which the welfare function is given by

$$W(e_i, E) = B(e_i) - D(E), \quad (23)$$

where $B(e_i)$ with $B' > 0$, $B'' < 0$ is the benefit of emissions and $D(E)$ with $D' > 0$, $D'' > 0$ is the environmental damage. There are no stability results for the general functional forms (23). In order to get specific results, Eckert and Nkuiya (2022) use the parametric functional forms

¹⁶ In Lemma 8 of the Appendix, we prove that Assumptions 1 and 3 are satisfied in the games of this section.

$$B(e_i) = ae_i - \frac{b}{2}e_i^2, \quad D(E) = \frac{d}{\alpha}E^\alpha, \quad (24)$$

where a , b and d are positive parameters, and¹⁷ $1 < \alpha \leq 2$ is the elasticity of the environmental damage with respect to global emissions.

Barrett (1994) was the first who introduced an abatement game in which the welfare function is given by

$$W(q_i, Q) = \frac{b}{n} \left(aQ - \frac{1}{2}Q^2 \right) - \frac{c}{2}q_i^2, \quad (25)$$

where $q_i = e^{\text{BAU}} - e_i$ is country i 's abatement compared to BAU, $Q := \sum_{j \in N} q_j = E^{\text{BAU}} - E$ is global abatement compared to BAU, a , b and c are positive parameters, and n is the number of countries. Diamantoudi and Sartzetakis (2006) and Rubio and Ulph (2006) show that the welfare functions (23) with (24) and (25) are transferable to one another for $\alpha = 2$.

In the Nash game with exogenous cooperation intensity $\gamma = 1$, Eckert and Nkuiya (2022) show by means of numerical simulations for $1 < \alpha \leq 2$ and Finus (2001) analytically proves for $\alpha = 2$ that the stable coalition consists of at most two countries.¹⁸ In the Stackelberg game with exogenous cooperation intensity $\gamma = 1$, Diamantoudi and Sartzetakis (2006) analytically find for¹⁹ $\alpha = 2$ that a coalition of at most four countries is stable, whereas Barrett (1994) with the help of simulations and Rubio and Ulph (2006) analytically show that any coalition can be stable.²⁰

Nkuiya (2020) uses the following isoelastic benefit and quadratic damage function

$$B(e_i) = \frac{e_i^{1-g} - 1}{1-g}, \quad D(E) = \frac{d}{2}E^2, \quad (26)$$

where d and g are positive parameters. Nkuiya (2020) finds that the stable coalition consists of at most two countries in the Nash game with exogenous cooperation intensity $\gamma = 1$, whereas in the Stackelberg game with exogenous cooperation intensity $\gamma = 1$ large coalitions including the grand coalition can be stable.²¹

4.2 Emissions and Trade

Eichner and Pethig (2013, 2015) investigate the impact of trade on the stability of climate coalitions. They develop a general equilibrium model with a composite consumer good and

¹⁷ Assumptions 2 and 3 are violated for $\alpha > 2$.

¹⁸ Eckert and Nkuiya (2022) show that large coalitions including the grand coalition can be stable in the Nash game with highest cooperation intensity if $\alpha > 2$.

¹⁹ There are no results for the Stackelberg game with $\alpha \neq 2$.

²⁰ Diamantoudi and Sartzetakis (2006) restrict their analysis to strictly positive emissions, Rubio and Ulph (2006) allow for corner solutions, and in Barrett (1994) emissions are unrestricted.

²¹ Finus et al. (2024) prove for generic benefit and damage functions that the stable coalition is larger in Stackelberg games with exogenous cooperation intensity $\gamma = 1$ than in Nash games with exogenous cooperation intensity $\gamma = 1$.

fossil fuel which are produced and consumed in each country and internationally traded. Fossil fuel consumption causes carbon emissions. Denoting by e_i country i 's emissions cap its welfare function is given by

$$W(e_i, E) = V(e_i) + T\left(\frac{E}{n}\right) - \left(\frac{E}{n} - e_i\right) T'\left(\frac{E}{n}\right) - D(E), \quad (27)$$

where

$$V(e_i) = ae_i - \frac{b}{2}e_i^2, \quad T\left(\frac{E}{n}\right) = \bar{x} - \frac{c}{2}\left(\frac{E}{n}\right)^2, \quad D(E) = \frac{1}{2}E^2 \quad (28)$$

and a , b , c and \bar{x} are positive parameters. In (27) and (28), $V(e_i)$ represents country i 's utility of fossil fuel consumption, $T\left(\frac{E}{n}\right)$ is country i 's transformation function between the consumer good and fossil fuel and $D(E)$ represents country i 's environmental damage. In the Nash game with exogenous cooperation intensity $\gamma = 1$ the stable coalition consists of at most two countries but in the Stackelberg game with exogenous cooperation intensity $\gamma = 1$ stable coalitions may be larger and can include up to (slightly more than) 50 % of all countries when the parameter space is constrained to ensure non-negative emissions. Note that (27) collapses to (24) for $c = 0$ and $\alpha = 2$, such that stable coalitions can be larger in the Stackelberg game when allowing for corner solutions (Rubio and Ulph 2006).

4.3 Mitigation and Adaptation

Bayramoglu et al. (2018) and Finus et al. (2021) analyze the stability of climate coalitions when countries choose mitigation and adaptation. They consider a three-stage game in which at stage 1 countries decide whether they participate in the coalition, at stage 2 country $i \in N$ chooses mitigation q_i , where $\sum_{j \in N} q_j =: Q$, and at stage 3 country $i \in N$ chooses adaptation x_i .²² The countries' welfare function is given by

$$\check{W}(q_i, x_i, Q) = \left(bQ - \frac{g}{2}Q^2\right) + x_i(a - fQ) - \frac{c}{2}q_i^2 - \frac{d}{2}x_i^2, \quad (29)$$

where a , b , c , d and g are positive parameters, and f is a parameter with $f^2 < d(c + n^2g)/n^2$. Solving the country's first order condition of adaptation $\check{W}_x = a - fQ - dx_i = 0$ at stage 3 with respect to x_i , and inserting into (29) yields²³

$$\check{W}\left(q_i, \frac{a - fQ}{d}, Q\right) = \left(bQ - \frac{g}{2}Q^2\right) - \frac{c}{2}q_i^2 + \frac{(a - fQ)^2}{2d} =: W(q_i, Q). \quad (30)$$

²² Since adaptation is a private good that does not cause any externalities, and since the costs of mitigation and adaptation are additively separable, each country chooses the same level of adaptation $x_i = \check{x}^i(Q)$. Substituting $x_i = \check{x}^i(Q)$ into the welfare function $\check{W}(q_i, x_i, Q)$ yields the identical welfare function $\check{W}(q_i, \check{x}^i(Q), Q) =: W(q_i, Q)$ to which our three-stage game can be applied. This procedure may not be applicable to other games with two actions.

²³ Since the results of section 3 neither rest on the interpretation of emissions nor on the signs of the partial derivatives, we can use q_i , Q and e_i , E interchangeably.

(30) implies that emissions are strategic substitutes [complements] if $W_{qQ} + W_{QQ} = f^2 - dg < [>]0$. If $f^2 - dg < 0$, the stable coalition consists of at most two countries in the Nash game with exogenous cooperation intensity $\gamma = 1$ (see Bayramoglu et al. 2018). Finus et al. (2021) show that the stable coalition may increase up to the grand coalition in the Stackelberg game with exogenous cooperation intensity $\gamma = 1$. If $f^2 - dg > 0$, a coalition of three countries is stable and the grand coalition can be stable, whereby the condition for the grand coalition to be stable is weaker in the Nash game with exogenous cooperation intensity $\gamma = 1$ than in the Stackelberg game with exogenous cooperation intensity $\gamma = 1$ (Finus et al. 2021).

4.4 Fishing

Pintassilgo and Lindroos (2008) and Long and Flaaten (2011) analyze the stability of fishing coalitions with logistic stock growth and linear cost functions. At the first stage, country $i \in N$ decides on its membership in the fishing coalition and at the second stage, it chooses its fishing effort e_i . The corresponding profit function is given by

$$\check{W}(e_i, X) = pqXe_i - ce_i, \quad (31)$$

where c , p and q are positive parameters and X is the fish stock. The steady-state relation between stock growth $G(X)$ and total fishing effort $H(X)$ is given by

$$G(X) = rX \left(1 - \frac{X}{K}\right) = qXE = H(X), \quad (32)$$

where r and K are positive parameters. Using (31) in (32) yields

$$\check{W} \left(e_i, K \left(1 - \frac{qE}{r} \right) \right) = pqK \left(1 - \frac{qE}{r} \right) e_i - ce_i =: W(e_i, E). \quad (33)$$

Long and Flaaten (2011) show that the stable coalition comprises slightly more than 50 % of all countries in the Stackelberg game with exogenous cooperation intensity $\gamma = 1$, whereas Pintassilgo and Lindroos (2008) prove that there is no stable coalition in the Nash game with exogenous cooperation intensity $\gamma = 1$.²⁴

4.5 Summary and Illustration of Results

Table 1 summarizes the possible stable coalition sizes in the presented Nash games with both exogenous cooperation intensity $\gamma = 1$ and with optimal cooperation intensity γ^* .²⁵ The associated stable coalitions are listed in the third and fourth columns of Table 1, respectively.

²⁴ Our Propositions 1–3 can also be applied to cartel stability (see, e.g., Shaffer (1995); Thoron (1998); Konishi and Lin (1999) and Zu et al. (2012)).

²⁵ Recall that according to Proposition 3 the stable coalition of the Nash game with optimal intensity coincides with the stable coalition of the Stackelberg game with exogenous intensity $\gamma = 1$ if strategies are strategic substitutes. As mentioned below Proposition 1 in case of strategic complements the stable coalition is identical in the Nash game with optimal intensity and in the Nash game with exogenous intensity $\gamma = 1$.

In the emissions and abatement games, in the emissions and trade game, in the mitigation and adaptation game with $f^2 < dg$ and in the fishing game, strategies are strategic substitutes. In view of the third and fourth columns of Table 1, in games with strategic substitutes the optimal cooperation intensity relative to the exogenous intensity $\gamma = 1$ (weakly) increases the coalition size, and a coalition of two or more countries is stable. The Nash equilibrium with optimal cooperation intensity Pareto dominates the business-as-usual equilibrium and there are Nash games in which optimizing the cooperation intensity can stabilize the grand coalition and implement the social optimum.²⁶ We conclude that optimizing the intensity of cooperation within the coalition in Nash games with strategic substitutes may help to enlarge the stable coalition compared to Nash games with exogenous cooperation intensity $\gamma = 1$.

In the mitigation and adaptation game with $f^2 > dg$ strategies are strategic complements. In that case the Nash coalition chooses the corner solution $\gamma^* = 1$ and the stable coalition coincides with the stable coalition of a Nash game with exogenous intensity $\gamma = 1$. Here, we conclude that optimizing the intensity of cooperation within the coalition does not change the stable coalition size and does not help to enhance the stable coalition compared to Nash games with exogenous cooperation intensity $\gamma = 1$.

Finally, we illustrate the results of optimizing the cooperation intensity in a Nash coalition with Barrett's (1994) canonical abatement game presented in subsection 4.1. Figure 1 shows the stable coalition size m^* (left-hand side figure) and the optimal cooperation intensity γ^* (right-hand side figure) in dependence of the abatement benefit parameter b .²⁷ The stable coalition size increases in b and the grand coalition becomes stable for $b \geq 75$.

Table 1 Possible coalition sizes in Nash games

Game	Sources	Cooperation intensity	
		Exogenous $\gamma = 1$	Optimal γ^*
<i>Emissions and abatement</i>			
– linear-quadratic specification	B94, F01, DS06, RU06, EN22	$m^* \leq 2$	$m^* \in [2, n]$
– isoelastic specification	N20	$m^* \leq 2$	$m^* \in [2, n]$
<i>Emissions and trade</i>	EP13, EP15	$m^* \leq 2$	$m^* \in [2, n/2]$
<i>Mitigation and adaptation</i>			
– strategic substitutes	BFJ18, FFR21	$m^* \leq 2$	$m^* \in [2, n]$
– strategic complements	BFJ18, FFR21	$m^* \in \{3, n\}$	$m^* \in \{3, n\}$
Fishing	PL08, LF11	$m^* = 0$	$m^* \approx n/2$

Note: B94: Barrett (1994), F01: Finus (2001), DS06: Diamantoudi and Sartzetakis (2006), RU06: Rubio and Ulph (2006), PL08: Pintassilgo and Lindroos (2008), LF11: Long and Flaaten (2011), EP13: Eichner and Pethig (2013), EP15: Eichner and Pethig (2015), BFJ18: Bayramoglu et al. (2018), N20: Nkuiya (2020), FFR21: Finus et al. (2021), EN22: Eckert and Nkuiya (2022)

²⁶ In the Nash game with exogenous intensity $\gamma = 1$, the grand coalition is never stable. Endogenizing γ changes the free-rider incentives and may ensue $\gamma^* = 1$ and $m^* = n$. To put it differently, although in the Nash game with exogenous intensity it holds $\gamma = 1$ and in the Nash game with optimizing the intensity it results $\gamma^* = 1$, the associated equilibrium emissions and the stable coalitions differ.

²⁷ The left-hand side figure shows the rational number m^* that solves the internal stability condition for the respective value of b . Note that the internal stability condition and the optimal cooperation intensity are functions of $b/c, m, n$, so that we can normalize $a, c = 1$ without loss of generality. See “Derivation of Figure 1” of the Appendix for further details.

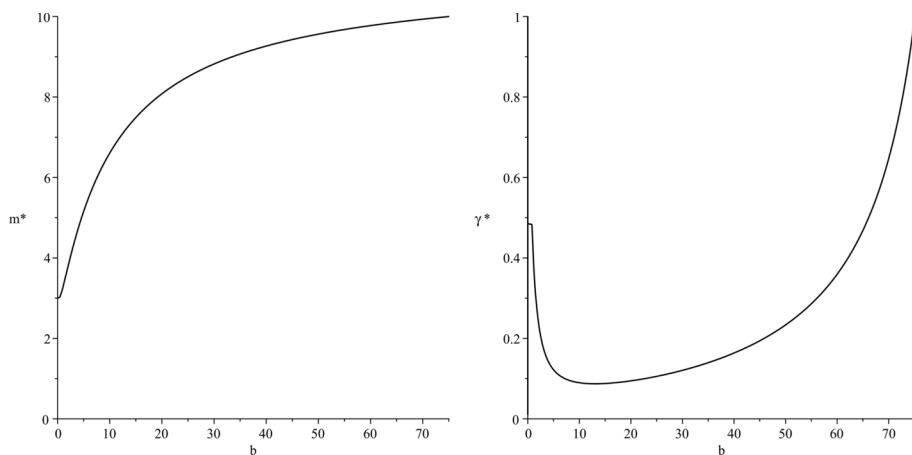


Fig. 1 Stable coalition size (left-hand side figure) and optimal cooperation intensity (right-hand side figure) dependent on b for $a, c = 1$ and $n = 10$ in Barrett's (1994) game with a Nash coalition

The optimal cooperation intensity is u-shaped in b and $\gamma^* = 1$ for $b \geq 75$. On the one hand, the optimal cooperation intensity *ceteris paribus* decreases with b because greater abatement benefits are accompanied by greater abatement efforts of non-members, which reduces the abatement incentives of the coalition members. On the other hand, the optimal cooperation intensity *ceteris paribus* increases with m because larger coalitions are accompanied by smaller leakage ($\partial \mathcal{R.F}' / \partial m > 0$). The right-hand side figure reveals that the former effect dominates for small coalitions and the latter effect dominates for large coalitions. However, no matter which of these effects dominates, the stable coalition size m^* increases in b . By contrast, with exogenous cooperation intensity $\gamma = 1$, a coalition of at most two countries is stable for any given b (see Table 1).

5 Conclusion

The present paper has analyzed the optimal intensity of cooperation in a general setting of self-enforcing environmental agreements. If signatories choose emissions before non-signatories (Stackelberg game), then there is no additional strategic advantage of setting a lower than the highest possible cooperation intensity. By contrast, if signatories and non-signatories choose emissions simultaneously (Nash game), then signatories set a weakly lower cooperation intensity than the highest cooperation intensity in case of strategic substitutes, and the highest cooperation intensity in case of strategic complements. In the former case, the subgame perfect equilibrium of the Nash game is identical to the subgame perfect equilibrium of the Stackelberg game.

The results of choosing the optimal cooperation intensity in the Nash game depend on whether strategies are strategic substitutes or complements and are applied to games in the literature. If emissions are strategic substitutes, optimizing with respect to the cooperation intensity may increase the stable coalition up to the grand coalition. By contrast, if emissions are strategic complements signatories choose the highest intensity of cooperation such that optimizing with respect to the cooperation intensity does not change the stable coali-

tion. In light of the results in games with strategic substitutes, our policy conclusion is that optimizing the intensity of cooperation may be helpful for enlarging the stable coalition. In view of the Paris Agreement, which seems to be a grand coalition with low intensity of cooperation, intensifying and optimizing the cooperation intensity may be worth trying to reach a grand coalition with the highest intensity of cooperation.

In Finus and Maus (2008) countries effectively agree on the level of modesty before the agreement is concluded, whereas in Harstad (2023b) the cooperation intensity results from exogenous uncertainty in a pledge-and-review bargaining game between the signatories. We contribute to this literature by letting the countries agree on the cooperation intensity after the agreement is concluded. Our model is therefore time consistent and uses the same cooperative solution concept for both the emissions choice and the cooperation intensity choice. Concerning the results the main difference is that in our approach the stable grand coalition with highest cooperation intensity may ensue whereas in Finus and Maus (2008) and Harstad (2023b) the stable grand coalition is characterized by a low or moderate cooperation intensity.

Our analysis can be extended in several directions. Although the optimal cooperation intensity (weakly) increases the coalition size in case of strategic substitutes, it is not clear whether this reduces global emissions or raises global welfare compared to the highest possible cooperation intensity. Furthermore, the analysis could be extended to the case of asymmetric countries with or without transfer payments. Finally, the analysis is restricted to models with one externality. The incorporation of positive spillovers from research and development of mitigation or adaptation technologies are on the agenda for future research in the field of self-enforcing environmental agreements with optimal cooperation intensity.

Appendix

Lemma 1

Under Assumption 1, at stage 3 of the Nash game there exists a unique equilibrium.

Proof

Following Cornes and Hartley (2007), Bayramoglu et al. (2018) and Finus et al. (2021, 2024), we use the concept of replacement functions to establish the existence and uniqueness of the equilibrium at stage 3 of the Nash game. Using $e_i = e_f \forall i \in F$ and $e_i = e_c \forall i \in C$ in (3) and (6), the equilibrium is characterized by

$$W_e^f(e_f, E) + W_E^f(e_f, E) = 0 \quad \forall i \in F, \quad (\text{A1})$$

$$W_e^c(e_c, E) + (1 + \gamma(m - 1))W_E^c(e_c, E) = 0 \quad \forall i \in C, \quad (\text{A2})$$

$$(n - m)e_f + me_c = E, \quad (\text{A3})$$

From these equations, e_f and e_c are functions of E , with the non-signatories' replacement function $e_f = \hat{R}^f(E)$, the signatories' replacement function $e_c = \hat{R}^c(E)$, and the aggregate replacement function $E = \hat{R}(E) = (n - m)\hat{R}^f(E) + m\hat{R}^c(E)$. Graphically, the equilibrium is given by the intersection of the aggregate replacement function with the 45-degree line. If the slope of the aggregate replacement function is less than one over the entire domain, it crosses the 45-degree line once, which implies the existence and uniqueness of the equilibrium. Totally differentiating (A1) and (A2) and rearranging yields the slopes of the individual replacement functions

$$\hat{R}_E^f = -\frac{W_{eE}^f + W_{EE}^f}{W_{ee}^f + W_{Ee}^f}, \quad (\text{A4})$$

$$\hat{R}_E^c = -\frac{W_{eE}^c + (1 + \gamma(m - 1))W_{EE}^c}{W_{ee}^c + (1 + \gamma(m - 1))W_{Ee}^c} \quad (\text{A5})$$

and the slope of the aggregate replacement function

$$\hat{R}_E = -(n - m)\frac{W_{eE}^f + W_{EE}^f}{W_{ee}^f + W_{Ee}^f} - m\frac{W_{eE}^c + (1 + \gamma(m - 1))W_{EE}^c}{W_{ee}^c + (1 + \gamma(m - 1))W_{Ee}^c} \quad (\text{A6})$$

$$\begin{aligned} &= 1 - (n - m)\frac{W_{ee}^f + W_{Ee}^f + n(W_{eE}^f + W_{EE}^f)}{n(W_{ee}^f + W_{Ee}^f)} \\ &\quad - m\frac{W_{ee}^c + (1 + \gamma(m - 1))W_{Ee}^c + n[W_{eE}^c + (1 + \gamma(m - 1))W_{EE}^c]}{n[W_{ee}^c + (1 + \gamma(m - 1))W_{Ee}^c]} \\ &= 1 - \frac{n - m}{n}\frac{(n - 1)(W_{ee}^f + nW_{Ee}^f) + W_{ee}^f + 2nW_{eE}^f + n^2W_{EE}^f}{n(W_{ee}^f + W_{Ee}^f)} \\ &\quad - \frac{m}{n}\frac{[(1 - \gamma)(n - 1) + \gamma(n - m)](W_{ee}^c + nW_{Ee}^c) + (1 + \gamma(m - 1))[W_{ee}^c + 2nW_{Ee}^c + n^2W_{EE}^c]}{n[W_{ee}^c + (1 + \gamma(m - 1))W_{Ee}^c]}. \end{aligned} \quad (\text{A7})$$

$W_{ee} < 0$, $W_{eE} \leq 0$ and $W_{ee} + 2nW_{eE} + n^2W_{EE} < 0$ from Assumption 1 implies that both fractions in (A7) are positive. Consequently, the slope of the aggregate replacement function is less than one over the entire domain, which is sufficient for the existence and uniqueness of the equilibrium at stage 3 of the Nash game. In the grand coalition equilibrium with the highest intensity of cooperation, i.e. $m = n$ and $\gamma = 1$, (A7) becomes

$$\hat{R}_E|_{m=n, \gamma=1} = 1 - \frac{W_{ee}^c + 2nW_{eE}^c + n^2W_{EE}^c}{W_{ee}^c + nW_{Ee}^c} = n - \frac{n(W_{ee}^c + nW_{eE}^c + W_{eE}^c + nW_{EE}^c)}{W_{ee}^c + nW_{Ee}^c}. \quad (\text{A8})$$

The second-order condition (7) requires $W_{ee}^c + nW_{eE}^c + W_{eE}^c + nW_{EE}^c < 0$. Consequently, $W_{ee}^c + nW_{Ee}^c < 0$ is necessary for $\hat{R}_E|_{m=n, \gamma=1} < n$, and $W_{ee}^c + 2nW_{eE}^c + n^2W_{EE}^c < 0$ is then necessary and sufficient for $\hat{R}_E|_{m=n, \gamma=1} < 1$. \square

Lemma 2

Under Assumption 1, at stage 2 of the Nash game it holds

$$\hat{e}_\gamma^f W_E^c (W_{eE}^f + W_{EE}^f) > 0, \quad \hat{e}_\gamma^c W_E^c > 0, \quad \hat{E}_\gamma W_E^c > 0, \quad (\hat{e}_\gamma^f - \hat{e}_\gamma^c) W_E^c < 0.$$

Proof

Differentiating (A1)-(A3) with respect to γ and solving for $\frac{de_f}{d\gamma}$, $\frac{de_c}{d\gamma}$ and $\frac{dE}{d\gamma}$ yields

$$\frac{de_f}{d\gamma} = \hat{e}_\gamma^f = \frac{nm(m-1)W_E^c(W_{eE}^f + W_{EE}^f)}{\Phi}, \quad (\text{A9})$$

$$\frac{de_c}{d\gamma} = \hat{e}_\gamma^c = -\frac{n(m-1)W_E^c[W_{ee}^f + W_{Ee}^f + (n-m)(W_{eE}^f + W_{EE}^f)]}{\Phi}, \quad (\text{A10})$$

$$\frac{dE}{d\gamma} = \hat{E}_\gamma = -\frac{nm(m-1)W_E^c(W_{ee}^f + W_{Ee}^f)}{\Phi} \quad (\text{A11})$$

and

$$\frac{d(e_f - e_c)}{d\gamma} = \hat{e}_\gamma^f - \hat{e}_\gamma^c = -\frac{n(m-1)W_E^c[W_{ee}^f + W_{Ee}^f + n(W_{eE}^f + W_{EE}^f)]}{\Phi}, \quad (\text{A12})$$

where

$$\begin{aligned} \Phi := & (n-m)[W_{ee}^c + (1+\gamma(m-1))W_{Ee}^c][W_{ee}^f + W_{Ee}^f + n(W_{eE}^f + W_{EE}^f)] \\ & + m(W_{ee}^f + W_{Ee}^f)[W_{ee}^c + (1+\gamma(m-1))W_{Ee}^c + n(W_{eE}^c + (1+\gamma(m-1))W_{EE}^c)] \end{aligned} \quad (\text{A13})$$

Along the lines of the proof of Lemma 1, Assumption 1 implies $\Phi > 0$ and, thus, $\hat{e}_\gamma^f W_E^c (W_{eE}^f + W_{EE}^f) > 0$, $\hat{e}_\gamma^c W_E^c > 0$, $\hat{E}_\gamma W_E^c > 0$ and $(\hat{e}_\gamma^f - \hat{e}_\gamma^c) W_E^c < 0$. \square

Lemma 3

Under Assumptions 1 and 2, at stage 2 of the Nash game the second-order condition (12) is satisfied.

Proof

Substituting (5) into (12) and rearranging yields

$$\frac{1}{n^2[W_{ee}^f + W_{Ee}^f + (n-m)(W_{ee}^f + W_{EE}^f)]^2} \cdot \left\{ (n-m)^2[W_{ee}^f + W_{Ee}^f + n(W_{ee}^f + W_{EE}^f)]^2(W_{ee}^c + m^2\mathcal{RF}''W_E^c) + 2m(n-m)(W_{ee}^f + W_{Ee}^f)[W_{ee}^f + W_{Ee}^f + n(W_{ee}^f + W_{EE}^f)] \cdot (W_{ee}^c + m^2\mathcal{RF}''W_E^c + nW_{eE}^c) + m^2(W_{ee}^f + W_{Ee}^f)^2[W_{ee}^c + m^2\mathcal{RF}'' \cdot W_E^c + nW_{eE}^c + n(W_{Ee}^c + nW_{EE}^c)] \right\} < 0, \quad (\text{A14})$$

such that second-order condition (12) is satisfied for $W_{ee} < 0$, $W_{ee} + nW_{eE} < 0$ and $W_{ee} + 2nW_{eE} + n^2W_{EE} < 0$ from Assumption 1 and $\mathcal{RF}'' \cdot W_E^c \leq 0$ from Assumption 2. \square

Lemma 4

Under Assumption 1, at stage 2 of the Nash game it holds

$$\begin{aligned} \hat{W}^f(m) > \hat{W}^c(m) > w^{\text{BAU}} &\iff G(m) > 0, \\ \hat{W}^f(m) = \hat{W}^c(m) = w^{\text{BAU}} &\iff G(m) = 0. \end{aligned}$$

Proof

We first prove that at stage 2 of the Nash game it holds

$$W_E^f W_E^c > 0, \quad W_E^c(W_E^f - W_E^c) \geq 0. \quad (\text{A15})$$

Note that Lemma 1 implies $e_f = e_c = e^{\text{BAU}}$ if $\gamma = 0$, such that $W_E^f = W_E^c$ if $\gamma = 0$. Now suppose $\gamma > 0$ and $W_E^c > [<]0$, such that the first-order condition (A2) implies $W_e^c < [>]0$, and Lemma 2 implies $e_c > [<]e_f$. Then, $W_{Ee} \leq 0$ implies $W_E^c \leq [>]W_E^f$ and, thus, $0 < [>]W_E^c \leq [>]W_E^f$ and $W_E^c(W_E^f - W_E^c) \geq 0$. This proves (A15).

Note that Lemma 1 implies $\hat{W}^f(m) = \hat{W}^c(m) = w^{\text{BAU}}$ if $\gamma = 0$. Differentiating W^f , W^c and $W^f - W^c$ with respect to γ , using the first-order conditions (A1) and (A2) as well as $\frac{de_f}{d\gamma}$, $\frac{de_c}{d\gamma}$ and $\frac{dE}{d\gamma}$ from (A9), (A10) and (A11), and rearranging yields

$$\begin{aligned}\frac{dW^f}{d\gamma} &= W_e^f \frac{de_f}{d\gamma} + W_E^f \frac{dE}{d\gamma} = W_E^f \left[\frac{dE}{d\gamma} - \frac{de_f}{d\gamma} \right] \\ &= -nm(m-1)W_E^c W_E^f \frac{W_{ee}^f + W_{Ee}^f + W_{eE}^f + W_{EE}^f}{\Phi},\end{aligned}\quad (\text{A16})$$

$$\begin{aligned}\frac{dW^c}{d\gamma} &= W_e^c \frac{de_c}{d\gamma} + W_E^c \frac{dE}{d\gamma} = W_E^c \left[\frac{dE}{d\gamma} - (1 + \gamma(m-1)) \frac{de_c}{d\gamma} \right] \\ &= -[\Gamma(m) - \gamma] n(m-1)^2 (W_E^c)^2 \frac{W_{ee}^f + W_{Ee}^f + (n-m)(W_{eE}^f + W_{EE}^f)}{\Phi},\end{aligned}\quad (\text{A17})$$

$$\begin{aligned}\frac{d(W^f - W^c)}{d\gamma} &= -nm(m-1)W_E^c (W_E^f - W_E^c) \frac{W_{ee}^f + W_{Ee}^f + W_{eE}^f + W_{EE}^f}{\Phi} \\ &\quad - n(m-1)(W_E^c)^2 \frac{W_{ee}^f + W_{Ee}^f + n(W_{eE}^f + W_{EE}^f)}{\Phi} \\ &\quad - \gamma n(m-1)^2 (W_E^c)^2 \frac{W_{ee}^f + W_{Ee}^f + (n-m)(W_{eE}^f + W_{EE}^f)}{\Phi}.\end{aligned}\quad (\text{A18})$$

Using (A15) in (A16) reveals $\frac{dW^f}{d\gamma} > 0$ and, thus, $W^f > [=]w^{\text{BAU}}$ for $G(m) > [=]0$.

(A17) reveals $\frac{dW^c}{d\gamma} \gtrless 0$ for $\gamma \gtrless \Gamma(m)$ and, thus, $W^c > [=]w^{\text{BAU}}$ for $G(m) > [=]0$.

Finally, using (A15) in (A18) reveals $\frac{d(W^f - W^c)}{d\gamma} > 0$ and, thus, $W^f > [=]W^c$ for $G(m) > [=]0$. \square

Proof of Proposition 1

We first prove that there does not exist a SPE with $\gamma^* = 0$. Suppose there exists a SPE with $G(m^*) = 0$. Then, $\hat{W}^f(m^*) = w^{\text{BAU}}$ and $\hat{W}^c(m) \geq w^{\text{BAU}}$ from Lemma 3 implies $\hat{W}^f(m^*) = w^{\text{BAU}} \leq \hat{W}^c(m^* + 1)$, such that m^* is externally unstable. Consequently, there does not exist a SPE with $\gamma^* = 0$.

Now we prove that there exists at least one SPE with $\gamma^* > 0$ and $m^* \geq 2$. Consider the grand coalition with $G(n) = 1$. If $m = n$ is internally stable, then it is an equilibrium coalition. Else, if it is not internally stable, then $m = n - 1$ is externally stable,²⁸ such that $G(n - 1) > 0$.²⁹ Again, either $m = n - 1$ is an equilibrium coalition, or $m = n - 2$ is externally stable with $G(n - 2) > 0$. Repeating this argument, either some coalition $m \geq 3$ is an equilibrium coalition, or $m = 2$ is externally stable with $G(2) > 0$. Then, $\hat{W}^f(1) = w^{\text{BAU}}$ and $\hat{W}^c(2) > w^{\text{BAU}}$ from Lemma 3 implies $\hat{W}^f(1) = w^{\text{BAU}} < \hat{W}^c(2)$, such that $m = 2$ is internally stable. Consequently, there exists at least one SPE with $\gamma^* > 0$ and $m^* \geq 2$.

If $m^* < n$ [$m^* = n$] and $W_{eE}^f + W_{EE}^f < 0$, such that emissions are strategic substitutes and $\mathcal{R}\mathcal{F}' \in]-1, 0[$ [$\mathcal{R}\mathcal{F}' = 0$] from (5), the SPE is characterized by $\gamma^* \in]0, 1[$ [$\gamma^* = 1$] from (13), which proves Proposition 1(i). If $W_{eE}^f + W_{EE}^f \geq 0$, such that emissions are stra-

²⁸ $\hat{W}^c(m) < \hat{W}^f(m - 1)$ implies $\hat{W}^f(m') > \hat{W}^c(m' + 1)$ with $m' = m - 1$.

²⁹ Otherwise the external stability condition would be violated.

tegic complements and $\mathcal{RF}' \geq 0$ from (5), the SPE is characterized by $\gamma^* = 1$ from (13), which proves Proposition 1(ii). \square

Lemma 5

Under Assumptions 1 and 2, at stage 3 of the Stackelberg game the second-order condition (18) is satisfied and there exists a unique equilibrium.

Proof

We first prove that the second-order condition (18) is satisfied. Substituting (5) into (18) and rearranging yields

$$\begin{aligned} & \frac{1}{n^2[W_{ee}^f + W_{Ee}^f + (n-m)(W_{eE}^f + W_{EE}^f)]^2} \\ & \cdot \left\{ (n-m)^2[W_{ee}^f + W_{Ee}^f + n(W_{eE}^f + W_{EE}^f)]^2(W_{ee}^c + \Xi) \right. \\ & \quad + 2m(n-m)(W_{ee}^f + W_{Ee}^f)[W_{ee}^f + W_{Ee}^f + n(W_{eE}^f + W_{EE}^f)] \\ & \quad \cdot \left[W_{ee}^c + \Xi + \frac{n}{2m}W_{eE}^c + \frac{n(1+\gamma(m-1))}{2m}W_{Ee}^c \right] \\ & \quad \left. + m^2(W_{ee}^f + W_{Ee}^f)^2 \left[W_{ee}^c + \Xi + \frac{n}{m}W_{eE}^c + \frac{n(1+\gamma(m-1))}{m} \left(W_{Ee}^c + \frac{n}{m}W_{EE}^c \right) \right] \right\} < 0, \end{aligned} \quad (\text{A19})$$

where $\Xi := (1 + \gamma(m-1))\mathcal{RF}'' \cdot W_E^c$, such that second-order condition (18) is satisfied for $W_{ee} < 0$, $W_{ee} + nW_{eE} < 0$ and $W_{ee} + 2nW_{eE} + n^2W_{EE} < 0$ from Assumption 1 and $\mathcal{RF}'' \cdot W_E^c \leq 0$ from Assumption 2.

Along the lines of the proof of Lemma 1, we use the concept of replacement functions to establish the existence and uniqueness of the equilibrium at stage 3 of the Stackelberg game. Using $e_i = e_f \forall i \in F$ and $e_i = e_c \forall i \in C$ in (3) and (17), the equilibrium is characterized by

$$W_e^f(e_f, E) + W_E^f(e_f, E) = 0 \quad \forall i \in F, \quad (\text{A20})$$

$$W_e^c(e_c, E) + (1 + \gamma(m-1))[1 + \mathcal{RF}'(E)]W_E^c(e_c, E) = 0 \quad \forall i \in C, \quad (\text{A21})$$

$$(n-m)e_f + me_c = E, \quad (\text{A22})$$

From these equations, e_f and e_c are functions of E , with the non-signatories' replacement function $e_f = \tilde{R}^f(E)$, the signatories' replacement function $e_c = \tilde{R}^c(E)$, and the aggregate replacement function $E = \tilde{R}(E) = (n-m)\tilde{R}^f(E) + m\tilde{R}^c(E)$. Totally differentiating (A20) and (A21) and rearranging yields the slopes of the individual replacement functions

$$\tilde{R}_E^f = -\frac{W_{eE}^f + W_{EE}^f}{W_{ee}^f + W_{Ee}^f}, \quad (\text{A23})$$

$$\tilde{R}_E^c = -\frac{W_{eE}^c + (1 + \gamma(m-1))[(1 + \mathcal{RF}')W_{EE}^c + \frac{\mathcal{RF}''}{1+\mathcal{RF}'}W_E^c]}{W_{ee}^c + (1 + \gamma(m-1))(1 + \mathcal{RF}')W_{Ee}^c}, \quad (\text{A24})$$

where $\frac{d\mathcal{RF}'}{dE} = \mathcal{RF}'' / \frac{dE}{dc} = \frac{\mathcal{RF}''}{1+\mathcal{RF}'}$, and the slope of the aggregate replacement function

$$\tilde{R}_E = -(n-m)\frac{W_{eE}^f + W_{EE}^f}{W_{ee}^f + W_{Ee}^f} - m\frac{W_{eE}^c + (1 + \gamma(m-1))[(1 + \mathcal{RF}')W_{EE}^c + \frac{\mathcal{RF}''}{1+\mathcal{RF}'}W_E^c]}{W_{ee}^c + (1 + \gamma(m-1))(1 + \mathcal{RF}')W_{Ee}^c} \quad (\text{A25})$$

$$\begin{aligned} &= 1 - \frac{1/[W_{ee}^c + (1 + \gamma(m-1))(1 + \mathcal{RF}')W_{Ee}^c]}{n^2(W_{ee}^f + W_{Ee}^f)[W_{ee}^f + W_{Ee}^f + (n-m)(W_{eE}^f + W_{EE}^f)]} \\ &\quad \cdot \left\{ (n-m)^2[W_{ee}^f + W_{Ee}^f + n(W_{eE}^f + W_{EE}^f)]^2(W_{ee}^c + m\Xi) \right. \\ &\quad + 2m(n-m)(W_{ee}^f + W_{Ee}^f)[W_{ee}^f + W_{Ee}^f + n(W_{eE}^f + W_{EE}^f)] \\ &\quad \cdot \left[W_{ee}^c + m\Xi + \frac{n}{2}W_{eE}^c + \frac{n(1 + \gamma(m-1))}{2m}W_{Ee}^c \right] \\ &\quad + m^2(W_{ee}^f + W_{Ee}^f)^2 \\ &\quad \cdot \left[W_{ee}^c + m\Xi + nW_{eE}^c + \frac{n(1 + \gamma(m-1))}{m} \left(W_{Ee}^c + nW_{EE}^c \right) \right] \Big\}, \end{aligned} \quad (\text{A26})$$

which is smaller than one for $W_{ee} < 0$, $W_{eE} \leq 0$ and $W_{ee} + 2nW_{eE} + n^2W_{EE} < 0$ from Assumption 1 and $\mathcal{RF}'' \cdot W_E^c \leq 0$ from Assumption 2. Consequently, Assumptions 1 and 2 are sufficient for the existence and uniqueness of the equilibrium at stage 3 of the Stackelberg game. \square

Lemma 6

Under Assumption 1, at stage 2 of the Stackelberg game it holds

$$\begin{aligned} \tilde{W}^f(m) &> \tilde{W}^c(m) > w^{\text{BAU}} && \iff m(1 + \mathcal{RF}') > 1, \\ \tilde{W}^f(m) &= \tilde{W}^c(m) = w^{\text{BAU}} && \iff m(1 + \mathcal{RF}') = 1, \\ \tilde{W}^c(m) &> w^{\text{BAU}} > \tilde{W}^f(m) && \iff m(1 + \mathcal{RF}') < 1. \end{aligned}$$

Proof

The proof follows along the lines of the proofs of Lemmas 2 and 4. Using $e_i = e_f \forall i \in F$ and $e_i = e_c \forall i \in C$ in (3) and (21), the equilibrium is characterized by

$$W_e^f(e_f, E) + W_E^f(e_f, E) = 0 \quad \forall i \in F, \quad (\text{A27})$$

$$W_e^c(e_c, E) + \alpha W_E^c(e_c, E) = 0 \quad \forall i \in C, \quad (\text{A28})$$

$$(n - m)e_f + me_c = E, \quad (\text{A29})$$

where $\alpha := m(1 + \mathcal{RF}') > 0$. Differentiating (A27)–(A29) with respect to α and solving for $\frac{de_f}{d\alpha}$, $\frac{de_c}{d\alpha}$ and $\frac{dE}{d\alpha}$ yields

$$\frac{de_f}{d\alpha} = \frac{nmW_E^c(W_{ee}^f + W_{EE}^f)}{\Psi}, \quad (\text{A30})$$

$$\frac{de_c}{d\alpha} = -\frac{nW_E^c[W_{ee}^f + W_{Ee}^f + (n - m)(W_{eE}^f + W_{EE}^f)]}{\Psi}, \quad (\text{A31})$$

$$\frac{dE}{d\alpha} = -\frac{nmW_E^c(W_{ee}^f + W_{Ee}^f)}{\Psi}, \quad (\text{A32})$$

and

$$\frac{d(e_f - e_c)}{d\alpha} = -\frac{nW_E^c[W_{ee}^f + W_{Ee}^f + n(W_{eE}^f + W_{EE}^f)]}{\Psi}, \quad (\text{A33})$$

where

$$\begin{aligned} \Psi &:= (n - m)(W_{ee}^c + \alpha W_{Ee}^c)[W_{ee}^f + W_{Ee}^f + n(W_{eE}^f + W_{EE}^f)] \\ &\quad + m(W_{ee}^f + W_{Ee}^f)[W_{ee}^c + \alpha W_{Ee}^c + n(W_{eE}^c + \alpha W_{EE}^c)] \\ &= \frac{1}{n[W_{ee}^f + W_{Ee}^f + (n - m)(W_{eE}^f + W_{EE}^f)]} \\ &\quad \cdot \left\{ (n - m)^2[W_{ee}^f + W_{Ee}^f + n(W_{eE}^f + W_{EE}^f)]^2 W_{ee}^c \right. \\ &\quad + 2m(n - m)(W_{ee}^f + W_{Ee}^f)[W_{ee}^f + W_{Ee}^f + n(W_{eE}^f + W_{EE}^f)](W_{ee}^c + nW_{Ee}^c) \\ &\quad \left. + m^2(W_{ee}^f + W_{Ee}^f)^2[W_{ee}^c + nW_{Ee}^c + n(W_{eE}^c + nW_{EE}^c)] \right\}. \end{aligned} \quad (\text{A34})$$

Assumption 1 implies $\Psi > 0$ and, thus, $\frac{de_f}{d\alpha} W_E^c(W_{eE}^f + W_{EE}^f) > 0$, $\frac{de_c}{d\alpha} W_E^c > 0$, $\frac{dE}{d\alpha} W_E^c > 0$ and $\frac{d(e_f - e_c)}{d\alpha} W_E^c < 0$.

Next we prove

$$W_E^f W_E^c > 0, \quad \alpha \geq 1 \implies W_E^c(W_E^f - W_E^c) \geq 0. \quad (\text{A35})$$

Note that Lemma 5 implies $e_f = e_c = e^{\text{BAU}}$ if $\alpha = 1$, such that $W_E^f = W_E^c$ if $\alpha = 1$. Now suppose $\alpha > 1$ and $W_E^c > [\leq] 0$, such that the first-order condition (A28) implies $W_e^c < [\geq] 0$, and (A33) implies $e_c > [\leq] e_f$. Then, $W_{ee} \leq 0$ implies $W_e^c \leq [\geq] W_E^f$ and, thus, $0 < [\geq] W_E^c \leq [\geq] W_E^f$ and $W_e^c(W_E^f - W_E^c) \geq 0$. Finally suppose $\alpha \in]0, 1[$ and $W_E^c > [\leq] 0$, such that the first-order condition (A28) implies $W_e^c < [\geq] 0$, and (A33) implies $e_c < [\geq] e_f$. Then, $W_{ee} \leq 0$ implies $W_e^c \geq [\leq] W_E^f$ and, thus, $0 > [\leq] W_e^c \geq [\leq] W_E^f$, such that the first-order condition (A27) implies $W_E^f > [\leq] 0$. This proves (A35).

Note that Lemma 5 implies $\tilde{W}^f(m) = \tilde{W}^c(m) = w^{\text{BAU}}$ if $\alpha = 1$. Differentiating W^f , W^c and $W^f - W^c$ with respect to α , using the first-order conditions (A27) and (A28) as well as $\frac{de_f}{d\alpha}$, $\frac{de_c}{d\alpha}$ and $\frac{dE}{d\alpha}$ from (A30), (A31) and (A32), and rearranging yields

$$\begin{aligned} \frac{dW^f}{d\alpha} &= W_e^f \frac{de_f}{d\alpha} + W_E^f \frac{dE}{d\alpha} = W_E^f \left[\frac{dE}{d\alpha} - \frac{de_f}{d\alpha} \right] \\ &= -nmW_E^c W_E^f \frac{W_{ee}^f + W_{Ee}^f + W_{eE}^f + W_{EE}^f}{\Psi}, \end{aligned} \quad (\text{A36})$$

$$\begin{aligned} \frac{dW^c}{d\alpha} &= W_e^c \frac{de_c}{d\alpha} + W_E^c \frac{dE}{d\alpha} = W_E^c \left[\frac{dE}{d\alpha} - \alpha \frac{de_c}{d\alpha} \right] \\ &= -[m(1 + \mathcal{RF}') - \alpha] n(W_E^c)^2 \frac{W_{ee}^f + W_{Ee}^f + (n-m)(W_{eE}^f + W_{EE}^f)}{\Psi}, \end{aligned} \quad (\text{A37})$$

$$\begin{aligned} \frac{d(W^f - W^c)}{d\alpha} &= -nmW_E^c(W_E^f - W_E^c) \frac{W_{ee}^f + W_{Ee}^f + W_{eE}^f + W_{EE}^f}{\Psi} \\ &\quad - n(W_E^c)^2 \frac{W_{ee}^f + W_{Ee}^f + n(W_{eE}^f + W_{EE}^f)}{\Psi} \\ &\quad - (\alpha - 1)n(W_E^c)^2 \frac{W_{ee}^f + W_{Ee}^f + (n-m)(W_{eE}^f + W_{EE}^f)}{\Psi}. \end{aligned} \quad (\text{A38})$$

Using (A34) in (A36) reveals $\frac{dW^f}{d\alpha} > 0$ and, thus, $W^f \gtrless w^{\text{BAU}}$ for $m(1 + \mathcal{RF}') \gtrless 1$.

(A37) reveals $\frac{dW^c}{d\alpha} \gtrless 0$ for $\alpha \gtrless m(1 + \mathcal{RF}')$ and, thus, $W^c > [=] w^{\text{BAU}}$ for

$m(1 + \mathcal{RF}') \neq [=] 1$. Consequently, $\frac{d(W^f - W^c)}{d\alpha} > 0$ for $\alpha \in]0, 1[$ and, thus, $W^c > w^{\text{BAU}} > W^f$ for $m(1 + \mathcal{RF}') < 1$. Finally, using (A34) in (A38) reveals $\frac{d(W^f - W^c)}{d\alpha} > 0$ for $\alpha \geq 1$ and, thus, $W^f > [=] W^c$ for $m(1 + \mathcal{RF}') > [=] 1$. \square

Proof of Proposition 3

We first prove that there does not exist a SPE with $m^*(1 + \mathcal{RF}') \leq 1$. Suppose there exists a SPE with $m^*(1 + \mathcal{RF}') \leq 1$. Then, $\tilde{W}^f(m^*) \leq w^{\text{BAU}}$ and $\tilde{W}^c(m^*) \geq w^{\text{BAU}}$ from Lemma 6 implies $\tilde{W}^f(m^*) \leq w^{\text{BAU}} \leq \tilde{W}^c(m^* + 1)$, such that m^* is externally unstable. Consequently, there does not exist a SPE with $m^*(1 + \mathcal{RF}') \leq 1$. The proof that there

exists at least one SPE with $m^*(1 + \mathcal{RF}') > 1$ and $m^* \geq 2$ follows along the lines of the proof of Proposition 1.

If $m^* < n$ and $W_{eE}^f + W_{EE}^f < 0$, such that emissions are strategic substitutes and $\mathcal{RF}' \in]-1, 0[$ from (5), the SPE is characterized by $m^*(1 + \mathcal{RF}') \in]1, m^*[$ from $m^*(1 + \mathcal{RF}') > 1$. Consequently, the SPE of the Nash game with optimal cooperation intensity coincides with the SPE of the Stackelberg game with highest cooperation intensity if emissions are strategic substitutes. If $m^* = n$ or $W_{eE}^f + W_{EE}^f = 0$, such that the grand coalition is stable or the non-signatories have dominant strategies, the SPE is characterized by $m^*(1 + \mathcal{RF}') = m^*$. Consequently, the SPE of the Nash game with optimal cooperation intensity coincides with the SPE of the Stackelberg game with highest cooperation intensity if the grand coalition is stable or the non-signatories have dominant strategies. If $m^* < n$ and $W_{eE}^f + W_{EE}^f > 0$, such that emissions are strategic complements and $\mathcal{RF}' \geq 0$ from (5), the SPE is characterized by $m^*(1 + \mathcal{RF}') > m^*$. Consequently, the SPE of the Nash game with optimal cooperation intensity differs from the SPE of the Stackelberg game with highest cooperation intensity if emissions are strategic complements. \square

Lemma 7

Under Assumptions 1 and 3, Assumption 2 is satisfied.

Proof

Using $\frac{de_f}{dE} = -\frac{W_{eE}^f + W_{EE}^f}{W_{ee}^f + W_{Ee}^f}$ and $\frac{dE}{de_c} = 1 + \mathcal{RF}'$ in \mathcal{RF}'' yields

$$\begin{aligned} \mathcal{RF}'' &= \frac{\partial \mathcal{RF}'}{\partial E} \frac{dE}{de_c} + \frac{\partial \mathcal{RF}'}{\partial e_f} \frac{de_f}{de_c} = \left(\frac{\partial \mathcal{RF}'}{\partial E} + \frac{\partial \mathcal{RF}'}{\partial e_f} \frac{de_f}{dE} \right) \frac{dE}{de_c} \\ &= \left(\frac{\partial \mathcal{RF}'}{\partial E} - \frac{\partial \mathcal{RF}'}{\partial e_f} \frac{W_{eE}^f + W_{EE}^f}{W_{ee}^f + W_{Ee}^f} \right) (1 + \mathcal{RF}'), \end{aligned} \quad (\text{A39})$$

where

$$\frac{\partial \mathcal{RF}'}{\partial E} = -\frac{(n-m) \left[(W_{eEE}^f + W_{EEE}^f) (W_{ee}^f + W_{Ee}^f) - (W_{eeE}^f + W_{eEE}^f) (W_{eE}^f + W_{EE}^f) \right]}{\left[W_{ee}^f + W_{Ee}^f + (n-m) (W_{eE}^f + W_{EE}^f) \right]^2}, \quad (\text{A40})$$

$$\frac{\partial \mathcal{RF}'}{\partial e_f} = -\frac{(n-m) \left[(W_{eEE}^f + W_{EEE}^f) (W_{ee}^f + W_{Ee}^f) - (W_{eeE}^f + W_{eEE}^f) (W_{eE}^f + W_{EE}^f) \right]}{\left[W_{ee}^f + W_{Ee}^f + (n-m) (W_{eE}^f + W_{EE}^f) \right]^2}, \quad (\text{A41})$$

such that

$$\begin{aligned} \mathcal{R}\mathcal{F}''W_E^c = & \frac{(n-m) \left(W_{eE}^f + W_{EE}^f\right)^2}{- \left[W_{ee}^f + W_{Ee}^f + (n-m) \left(W_{eE}^f + W_{EE}^f\right)\right]^3} \\ & \cdot \left\{ \left(W_{EEE}^f + W_{eEE}^f\right) \left(\frac{W_{ee}^f + W_{Ee}^f}{W_{eE}^f + W_{EE}^f}\right)^2 - 2 \left(W_{eeE}^f + W_{eEE}^f\right) \frac{W_{ee}^f + W_{Ee}^f}{W_{eE}^f + W_{EE}^f} \right. \\ & \left. + W_{eee}^f + W_{eeE}^f \right\} W_E^c. \end{aligned} \quad (\text{A42})$$

$W_{ee} < 0$, $W_{eE} \leq 0$ and $W_{ee} + 2nW_{eE} + n^2W_{EE} < 0$ from Assumption 1 implies that $\mathcal{R}\mathcal{F}''W_E^c$ is weakly negative if $W_E^c W_{EEE}^f$, $W_E^c W_{eee}^f \leq 0$ and W_{eeE}^f , $W_{eEE}^f = 0$. \square

Lemma 8

The welfare and profit functions of section 4 satisfy Assumptions 1 and 3.

Proof

For (24), we have $W_{ee} = -b < 0$, $W_{eE} = 0$, $W_{ee} + 2nW_{eE} + n^2W_{EE} = -b - n^2d(\alpha - 1)E^{\alpha-2} < 0$, $W_E W_{EEE} = -d^2(\alpha - 1)(2 - \alpha)E^{2\alpha-4} \leq 0 \Leftrightarrow \alpha \in [1, 2]$ and $W_{eee}, W_{eeE}, W_{eEE} = 0$. For (25), we have $W_{ee} = -c < 0$, $W_{eE} = 0$, $W_{ee} + 2nW_{eE} + n^2W_{EE} = -c - nb < 0$ and zero third derivatives. For (26), we have $W_{ee} = -ge_i^{-(1+g)} < 0$, $W_{eE} = 0$, $W_{ee} + 2nW_{eE} + n^2W_{EE} = -ge_i^{-(1+g)} - n^2d < 0$, $W_E W_{eee} = -dg(1 + g)e_i^{-(2+g)}E < 0$ and $W_{eeE}, W_{eEE}, W_{EEE} = 0$. For (27), we have $W_{ee} = -b < 0$, $W_{eE} = -c/n$, $W_{ee} + 2nW_{eE} + n^2W_{EE} = -b - 2c + c - n^2 < 0$ and zero third derivatives. For (30), we have $W_{qq} = -c$, $W_{qQ} = 0$, $W_{qq} + 2nW_{qQ} + n^2W_{QQ} = -c - n^2(dg - f^2)/d < 0 \Leftrightarrow f^2 < d(c + n^2g)/n^2$ and zero third derivatives. For (33), we have $W_{ee} = 0$, $W_{eE} = -pq^2K/r < 0$, $W_{ee} + 2nW_{eE} + n^2W_{EE} = -2npq^2K/r < 0$ and zero third derivatives. Thus, the welfare and profit functions of section 4 satisfy Assumptions 1 and 3. \square

Derivation of Figure 1:

At stage 3, using (25) in (3), (6) and $Q = mq_c + (n - m)q_f$ yields the Nash equilibrium of Barrett's (1994) abatement subgame with a Nash coalition

$$\frac{b}{n}(a - Q) - cq_f = 0 \quad \Leftrightarrow \quad q_f = \frac{b}{nc}(a - Q), \quad (\text{A43})$$

$$[1 + \gamma(m - 1)]\frac{b}{n}(a - Q) - cq_c = 0 \quad \Leftrightarrow \quad q_c = [1 + \gamma(m - 1)]\frac{b}{nc}(a - Q), \quad (\text{A44})$$

$$Q = mq_c + (n - m)q_f \quad \Leftrightarrow \quad Q = \frac{ab[n + \gamma m(m - 1)]}{[n + \gamma m(m - 1)]b + nc}. \quad (\text{A45})$$

Furthermore, using (25) in (5) yields the slope of the non-signatories' aggregate best-response function

$$\mathcal{RF}' = -\frac{(n-m)b}{(n-m)b+nc} \in]-1, 0[. \quad (\text{A46})$$

At stage 2, using (A46) in $\Gamma(m) := 1 + \frac{m}{m-1} \mathcal{RF}'$ yields

$$\Gamma(m) = \frac{n(m-1)c - (n-m)b}{(m-1)[(n-m)b+nc]} \gtrless 0 \Leftrightarrow m \gtrless \frac{n(b+c)}{b+nc}, \quad (\text{A47})$$

where $\frac{\partial \Gamma(m)}{\partial m} = \frac{b[(n-m)^2b+n(n-2m+m^2)c]}{(m-1)^2[(n-m)b+nc]^2} > 0$ and $\frac{\partial \Gamma(m)}{\partial b} = -\frac{b}{c} \frac{\partial \Gamma(m)}{\partial c} = -\frac{nm(n-m)c}{(m-1)[(n-m)b+nc]^2} < 0$. Using (A47) in (13) yields

$$\gamma = G(m) := \begin{cases} \frac{n(m-1)c - (n-m)b}{(m-1)[(n-m)b+nc]} & \text{if } m > \frac{n(b+c)}{b+nc}, \\ 0 & \text{if } m \leq \frac{n(b+c)}{b+nc}. \end{cases} \quad (\text{A48})$$

At stage 3, the external stability condition is violated if $m \leq \frac{n(b+c)}{b+nc}$ (see paragraph above Proposition 1). If $m > \frac{n(b+c)}{b+nc}$, using (A43)-(A45) and (A48) in (25) yields

$$W^f(m) = w^{\text{BAU}} + \frac{ma^2b^2c(b+nc)[n(m-1)c - (n-m)b][(n-m)(2n-m)b^2 + n(4n-3m+m^2)bc + 2n^2c^2]}{2n^2(b+c)^2[(n-m)^2b^2 + n(2n-2m+m^2)bc + n^2c^2]^2}. \quad (\text{A49})$$

$$W^c(m) = w^{\text{BAU}} + \frac{a^2b^2c[n(m-1)c - (n-m)b]^2}{2n^2(b+c)^2[(n-m)^2b^2 + n(2n-2m+m^2)bc + n^2c^2]}, \quad (\text{A50})$$

where $w^{\text{BAU}} = \frac{a^2b^2[nb+(2n-1)c]}{2n^2(b+c)^2}$. Using (A49) and (A50) yields

$$W^c(m) - W^f(m-1) = \frac{a^2b^4c^3[(n-m)^2b^2 + n(2n-2m+m^2)bc + n^2c^2]^{-1}}{2[(n-m+1)^2b^2 + n(2n-4m+3+m^2)bc + n^2c^2]^2} \cdot \Omega(n, m, d), \quad (\text{A51})$$

where $d := \frac{b}{c}$ and where

$$\begin{aligned} \Omega(n, m, d) := & (n-m)^2(n-m+1)^2d^2 + n(n-m+1)[2(n-m)^2 + (m^2-1)(n-m+1)]d \\ & - n^2[(m^2-4m+2)(n-m)^2 - 8(m-1)(n-m) - (m-1)(m+5)] \\ & - n^3(m-1)[2(m-3)n + (m-2)(m^2-3m+4)]d^{-1} - n^4(m-1)(m-3)d^{-2}. \end{aligned} \quad (\text{A52})$$

Thus, $W^c(m) - W^f(m-1) \gtrless 0 \Leftrightarrow \Omega(n, m, d) \gtrless 0$ and $\frac{\partial \Omega(n, m, d)}{\partial d} > 0$. The left-hand side of Figure 1 implicitly plots $\Omega(n, m^*, d) = 0$ for $n = 10$ and $d = b \Leftrightarrow c = 1$. Since $\Omega(n, m^*, d) = 0$ cannot be explicitly solved for m^* , we solve $\gamma^* = \Gamma(m^*)$ for m^* , which yields

$$m^* = -\frac{[1 - \gamma^*(n-1) + n(1-\gamma^*)d^{-1}]}{2\gamma^*} + \sqrt{\left[\frac{[1 - \gamma^*(n-1) + n(1-\gamma^*)d^{-1}]}{2\gamma^*}\right]^2 + \frac{n(1-\gamma^*)(1+d^{-1})}{\gamma^*}}, \quad (\text{A53})$$

and use this expression in $\Omega(n, m^*, d) = 0$, which yields $\Omega(n, \gamma^*, d) = 0$. The right-hand side of Figure 1 implicitly plots $\Omega(n, \gamma^*, d) = 0$ for $n = 10$ and $d = b \Leftrightarrow c = 1$.

Funding Open Access funding enabled and organized by Projekt DEAL.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. We have not received any financial support for the conduct of the research and/or preparation of the article from any third party.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Altamirano-Cabrera J-C, Finus M, Dellink R (2008) Do abatement quotas lead to more successful climate coalitions? *The Manchester School* 76:104–129
- Barrett S (1994) Self-enforcing international environmental agreements. *Oxford Economic Papers* 46:878–894
- Barrett S (2002) Consensus treaties. *Journal of Institutional and Theoretical Economics (JITE) / Zeitschrift für die gesamte Staatswissenschaft* 158, 529–547
- Bayramoglu B, Finus M, Jacques JF (2018) Climate agreements in a mitigation-adaptation game. *Journal of Public Economics* 165:101–113
- Carraro C, Siniscalco D (1993) Strategies for the international protection of the environment. *Journal of Public Economics* 52:309–328
- Colombo L, Labrecciosa P, Van Long N (2022) A dynamic analysis of international environmental agreements under partial cooperation. *European Economic Review* 143:104036
- Cornes RC, Hartley R (2007) Aggregative public good games. *Journal of Public Economic Theory* 9:201–219
- d'Aspremont C, Jacquemin A, Gabszewicz JJ, Weymark J (1983) On the stability of collusive price leadership. *Canadian Journal of Economics* 16:17–25
- Diamantoudi E, Sartzetakis E (2006) Stable international environmental agreements: An analytical approach. *Journal of Public Economic Theory* 8:247–263
- Dimitrov R, Hovi J, Sprinz D, Sælen H, Underdal A (2019) Institutional and environmental effectiveness: Will the Paris Agreement work? *Wiley Interdisciplinary Reviews: Climate Change* 10:e583
- Eckert H, Nkuiya B (2022) Damage sensitivity and stability in international environmental agreements. *Oxford Economic Papers* 74:1063–1076
- Eichner T, Pethig R (2013) Self-enforcing environmental agreements and international trade. *Journal of Public Economics* 102:37–50
- Eichner T, Pethig R (2015) Is trade liberalization conducive to the formation of climate coalitions? *International Tax and Public Finance* 22:932–955
- Endres A (1997) Negotiating a climate convention: The role of prices and quantities. *International Review of Law and Economics* 17:147–156

- Falkner R (2016) The Paris Agreement and the new logic of international climate politics. *International Affairs* 92:1107–1125
- Finus M (2001) *Game theory and international environmental cooperation*. Edward Elgar, Cheltenham and Northampton
- Finus M, Furini F, Rohrer AV (2021) The efficacy of international environmental agreements when adaptation matters: Nash-Cournot vs Stackelberg leadership. *Journal of Environmental Economics and Management* 109:102461
- Finus M, Furini F, Rohrer AV (2024) The Stackelberg vs. Nash-Cournot Folk-theorem in international environmental agreements. *Economics Letters* 234:111481
- Finus M, Maus S (2008) Modesty may pay. *Journal of Public Economic Theory* 10:801–826
- Finus M, Rundshagen B (1998) Toward a positive theory of coalition formation and endogenous instrumental choice in global pollution control. *Public Choice* 96:145–186
- Harstad B (2023a) Pledge-and-review bargaining. *Journal of Economic Theory* 207:105574
- Harstad B (2023b) Pledge-and-review bargaining: From Kyoto to Paris. *Economic Journal* 133:1181–1216
- Hoel M (1992) International environmental conventions: The case of uniform reductions of emissions. *Environmental and Resource Economics* 2:141–159
- Konish H, Lin P (1999) Stable cartels with a Cournot fringe in a symmetric oligopoly. *Keio Economics Studies* 36:1–10
- Long LK, Flaaten O (2011) A Stackelberg analysis of the potential for cooperation in straddling stock fisheries. *Marine Resource Economics* 26:119–139
- Nkuiya B (2020) Stability of international environmental agreements under isoelastic utility. *Resource and Energy Economics* 59:101–128
- Pintassilgo P, Lindroos M (2008) Coalition formation in straddling stock fisheries: A partition function approach. *International Game Theory Review* 10:303–317
- Rogelj J, den Elzen M, Höhne N, Fransen T, Fekete H, Winkler H, Schaeffer R, Sha F, Riahi K, Meinshausen M (2016) Paris Agreement climate proposals need a boost to keep warming well below 2 degree Celsius. *Nature Climate Change* 534:631–639
- Rubio SJ, Ulph A (2006) Self-enforcing international environmental agreements revisited. *Oxford Economic Papers* 58:233–263
- Shaffer S (1995) Stable cartels with a Cournot fringe. *Southern Economic Journal* 61:744–754
- Thoron S (1998) Formation of a coalition-proof stable cartel. *Canadian Journal of Economics* 31:63–76
- UNFCCC (2015): Decision 1/CP. 21, Adoption of the Paris Agreement. Report of the Conference of the Parties on Its Twenty-First Session, Held in Paris from 30 November to 13 December 2015. Addendum Part Two: Action Taken by the Conference of the Parties at Its Twenty-First Session (FCCC/CP/2015/10/Add. 1)
- Zu L, Zhang J, Wang S (2012) The size of stable cartels: An analytical approach. *International Journal of Industrial Organization* 30:217–222

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.