

Eichner, Thomas; Schopf, Mark

Conference Paper

Pledge and Review Bargaining in Environmental Agreements: Kyoto vs. Paris

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2021: Climate Economics

Provided in Cooperation with:

Verein für Socialpolitik / German Economic Association

Suggested Citation: Eichner, Thomas; Schopf, Mark (2021) : Pledge and Review Bargaining in Environmental Agreements: Kyoto vs. Paris, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2021: Climate Economics, ZBW - Leibniz Information Centre for Economics, Kiel, Hamburg

This Version is available at:

<https://hdl.handle.net/10419/242450>

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.

Self-enforcing climate agreements: Kyoto versus Paris*

Thomas Eichner

Department of Economics, University of Hagen

Mark Schopf

Department of Economics, University of Hagen

Abstract

This paper compares the Kyoto Protocol and the Paris Agreement in the dynamic game of Battaglini and Harstad (2016). The asymmetric Nash solution of this game reflects the Paris Agreement, whereas the symmetric Nash solution reflects the Kyoto Protocol. In a large set of economies, the Kyoto Protocol is Pareto superior to the Paris Agreement. Although the stable climate coalition is large in the Paris Agreement and small in the Kyoto Protocol, the emissions reductions of a single coalition country are much more pronounced in the Kyoto Protocol, so that this per-country-emissions-reduction effect outweighs the disadvantageous coalition-size effect.

JEL classification: C71, F55, Q54

Key words: pledge and review, emissions, investments, stable coalition

*Eichner: Department of Economics, University of Hagen, Universitätsstr. 41, 58097 Hagen, Germany, email: thomas.eichner@fernuni-hagen.de; Schopf: Corresponding author, Department of Economics, University of Hagen, Universitätsstr. 41, 58097 Hagen, Germany, email: mark.schopf@fernuni-hagen.de.

1 Introduction

Climate change is one of the greatest challenges facing humanity over the next decades. Both the Kyoto Protocol and the Paris Agreement have been negotiated to stabilize the world climate at safe levels. In the Kyoto Protocol, the Annex I countries committed themselves to reduce their emissions by as much as 18% below 1990 levels until 2020. After problems of further developing the Kyoto Protocol at the 2009 Copenhagen conference, countries switched 2015 from the Kyoto Protocol to the Paris Agreement. Although the Paris Agreement articulates a long-term goal of keeping the increase in global average temperature to 1.5° Celsius, the current commitments to emissions reductions made by the signatories are not even sufficient to limit global warming to 2° Celsius (Hulme 2016, Rogelj 2016). That raises the question what type of agreement is more appropriate to bring down world emissions to safe levels.

The Kyoto Protocol and the Paris Agreement exhibit a number of differences. The most fundamental difference is that the Kyoto Protocol is a ‘top-down’-approach which is managed by a multilateral institution and which legally binds countries to meet their emissions-reduction commitments. In contrast, the Paris Agreement is a ‘bottom-up’-approach in which countries submit their Intended Nationally Determined Contributions (pledges) to reduce carbon emissions. Pledges are voluntary commitments. While only 37 countries ratified the Kyoto Protocol, the Paris Agreement has been signed or implemented by 195 nations since the 21st Conference of Parties in Paris.¹ Although there is almost universal state participation in the Paris Agreement, it is not clear whether it leads to stronger reductions of world emissions. Pessimists believe that the Paris Agreement does not represent a breakthrough but instead incrementally extends the business as usual (Bang et al. 2016).

There is a large literature that analyzes self-enforcing climate agreements and explicitly refers to the Kyoto Protocol. The prevailing approach is to apply the *symmetric* Nash

¹The Holy Sea cannot accede because it is no member of the UNFCCC, and until today Lybia neither has ratified the Paris Agreement nor made any pledges.

(bargaining) solution in coalition formation games in which the climate coalition is both internally and externally stable. In a basic static game, in which countries choose their efforts to reduce emissions at the second stage and decide whether to join the climate coalition at the first stage, the size of stable coalitions is not larger than four (Hoel 1992, Carraro and Siniscalco 1993, Barrett 1994, Rubio and Ulph 2006).² Rubio and Casino (2005) and Rubio and Ulph (2007) extend the basic static game to a dynamic setting with a pollution stock. Rubio and Casino (2005) show that the stable coalition remains small if countries decide only once whether or not to participate in a climate coalition. Rubio and Ulph (2007) consider a variable membership decision and find that the stable coalition may be large if the potential gains from cooperation are small.³ In the dynamic game of Battaglini and Harstad (2016), countries decide to participate in an international climate agreement, to choose emissions, investments in clean technologies and the duration of the agreement. If climate contracts are incomplete, the stable coalition may be large up to the grand coalition. The driving force for large stable coalitions is a hold-up problem. In the last period of the contract, coalition countries invest too little in order to improve their bargaining power in future negotiations. All these contributions have in common that coalition countries maximize the sum of their welfares, which is tantamount to applying the symmetric Nash solution.

Only two recent contributions analyze the impact of the Paris Agreement on emissions, investments and welfare theoretically. Caparrós (2020) applies a partial commitment bargaining model to determine countries' emissions pledges. With appropriate transfers, the mechanism implements the first-best solution in the short term. Introducing investments in abatement and assuming that these investments are not part of the contract, a hold-up

²The basic static model is refined in various directions. E.g. Barrett (2006) and Hoel and de Zeeuw (2010) analyzed R&D in breakthrough technologies, Bayramoglu et al. (2018) mitigation and adaptation, McEvoy and McGinty (2018) emissions taxes and Kornek and Edenhofer (2020) compensation funds.

³Kováč and Schmidt (2021) analyze a dynamic abatement game in which a long-term contract is suspended for one period if some country participates in climate negotiations but does not sign the climate contract. This delay of the long-term contract reduces the free-riding incentives and enlarges the stable coalition. Karp and Sakamoto (2021) introduce uncertainty about the outcome if climate negotiations fail in a dynamic abatement game and show the existence of multiple equilibria. The uncertainty about the outcome reduces [enhances] the stability of small [large] coalitions and enlarges the stable coalition in the long-run.

problem emerges and countries underinvest. However, Caparrós (2020) does not investigate whether the climate contract is self-enforcing.

Harstad (2021) develops a novel Pledge-and-Review bargaining game in which countries simultaneously propose pledges. A contract is concluded if no country finds the vector of pledges unacceptable. Harstad (2021) shows that the solution of the Pledge-and-Review bargaining game⁴ maximizes the *asymmetric* Nash product. Building on this approach, Harstad (2020) investigates different Nash solutions in a dynamic coalition formation game in which countries decide on emissions pledges and investments in green energy. Thereby, the asymmetric Nash solution with low [high] welfare weights is interpreted as Paris Agreement [Kyoto Protocol].⁵ The symmetric Nash solution is the limiting case with the largest possible welfare weights. In the asymmetric Nash solution with low welfare weights stable climate coalitions are large but the emissions reductions of a single coalition country are small. In contrast, in the symmetric Nash solution stable climate coalitions are small (three) but a single coalition country undertakes large emissions reductions. In view of total emissions, there is a coalition-size effect and a countervailing per-country-emissions-reduction effect when moving from an asymmetric to the symmetric Nash solution. Comparing the Paris Agreement and the Kyoto Protocol, the coalition-size effect overcompensates the per-country-emissions-reduction effect, so that total emissions are smaller and countries' welfare is higher in the Paris Agreement (Harstad 2020, Corollary 1 and Proposition 3).⁶

The present paper compares the Paris Agreement (asymmetric Nash solution) and the Kyoto Protocol (symmetric Nash solution) in the dynamic game of Battaglini and Harstad (2016). There are three differences between Battaglini and Harstad's (2016) model, which is applied in the present paper, and Harstad's (2020) model: First, in Harstad (2020) the coalition members either expect never to cooperate again or they expect the current contract

⁴A simplified version of the Pledge-and-Review bargaining game has been tested in the laboratory by Lippert and Tremewan (2021).

⁵The welfare weights correspond to the Nash bargaining parameters of the other coalition countries in the welfare function of a single coalition country.

⁶Except there are some exogenously given minimum participation rules.

to be superseded by an identical contract once the current contract expires. By contrast, in Battaglini and Harstad (2016) the coalition members expect the equilibrium contract to emerge once the current contract expires. Consequently, members of small coalitions prefer shorter contracts to admit new members sooner, which amplifies the hold-up problem and stabilizes larger equilibrium coalitions.⁷ Second, the investment cost function in green technology depends on both investments and the technology stock instead of just investments. Third, the coalition members choose time-dependent pledges instead of constant pledges for the entire contract duration.⁸ Comparing the Paris Agreement and the Kyoto Protocol in the dynamic game of Battaglini and Harstad (2016), we show in line with Harstad (2020) that there is a coalition-size-effect and a per-country-emissions-reduction effect. Restricting our attention to economies in which the stable coalition comprises 195 countries in case of the asymmetric Nash solution and 37 countries in case of the symmetric Nash solution, the per-country-emissions-reduction effect dominates the coalition-size-effect for a large set of economies, so that world emissions are lower and countries' welfare is higher in the latter case than in the former. That is, in contrast to Harstad (2020) there exists a large set of economies in which the Kyoto Protocol is Pareto superior to the Paris Agreement.

The remainder of the paper is organized as follows: In Section 2 the building blocks of the model are presented. In Section 3 we analyze the dynamic coalition formation game and characterize equilibrium pledges, investments, contract durations and stable coalitions for different Nash solutions. In Section 4 we characterize the economies in which the Kyoto Protocol with 37 countries is Pareto superior to the Paris Agreement with 195 countries and vice versa. Section 5 concludes.

⁷In contrast to Harstad (2020), Battaglini and Harstad (2016) can thereby explain equilibrium coalitions of more than three countries in the symmetric Nash solution without relying on minimum participation rules.

⁸Only the combination of stock-independent investment costs and time-independent pledges as in Harstad (2020) can explain the equilibrium contract duration to be finite. For stock-independent investment costs and time-dependent pledges we refer to Harstad (2020, p. 30). The case of stock-dependent investment costs and time-independent pledges is dealt with in our Online Appendix.

2 The model

The world economy consists of n countries. In each period $t \geq 1$ country $i \in N = \{1, \dots, n\}$ consumes energy that composes of fossil fuel energy $g_{i,t}$ and green energy $R_{i,t}$. The benefit of energy consumption is

$$B_i(g_{i,t}, R_{i,t}) = -\frac{b}{2}(\bar{y}_i - g_{i,t} - R_{i,t})^2, \quad (1)$$

where \bar{y}_i is an exogenously given satiation point and b is an energy preference parameter. Emission units are chosen such that $g_{i,t}$ denotes both fossil fuel energy consumption and carbon emissions from burning fuel by country i .

The stock of pollution evolves according to

$$G_t = q_G G_{t-1} + \sum_{j \in N} g_{j,t}, \quad (2)$$

where $1 - q_G \in [0, 1]$ is the natural depreciation rate. The climate damage from the emissions stock G_t is given by cG_t , where c is the marginal damage.

Green energy is produced by means of a green technology $R_{i,t}$. For sake of simplicity, the generation of green energy is proportional to green technology, whose stock increases with investments $r_{i,t}$, and evolves in time according to

$$R_{i,t+1} = q_R R_{i,t} + r_{i,t}. \quad (3)$$

In (3), $1 - q_R \in [0, 1]$ is the technological depreciation rate. Following Battaglini and Harstad (2016, p. 167) the investment cost function κ depends on investments $r_{i,t}$ and on technology stock $R_{i,t}$ according to

$$\kappa(r_{i,t}, R_{i,t}) = \frac{k}{2}(r_{i,t}^2 + 2q_R r_{i,t} R_{i,t}), \quad (4)$$

where k is a positive parameter. Making use of (3) in (4), the investment cost function can be written as $\kappa(\cdot) = \frac{k}{2}(R_{i,t+1}^2 - q_R^2 R_{i,t}^2)$.

The time from one consumption decision to the next is $\Delta > 0$, and the time from the investment decision to the availability of the developed technology is $\Lambda \in (0, \Delta]$. Then, the utility of country i in period t is given by

$$u_{i,t} = -\frac{b}{2} (\bar{y}_i - g_{i,t} - R_{i,t})^2 - cG_t - \frac{k}{2} \left(R_{i,t+1}^2 - q_R^2 R_{i,t}^2 \right) e^{-\rho(\Delta-\Lambda)}, \quad (5)$$

where $\rho > 0$ is the discount rate. The present value of current and future utility is given by $v_{i,t} = \sum_{\tau=t} \delta^{\tau-t} u_{i,\tau}$, where $\delta \equiv e^{-\rho\Delta} \in (0, 1)$ is the discount factor.

The present value $v_{i,t}$ can be represented by a discounted utility stream that depends on the two choice variables ($d_{i,t}$ and $R_{i,t+1}$) and is independent of past stock variables ($G_{t-\tau}$ and $R_{i,t-\tau}$ with $\tau \geq 1$):

Lemma 1 (Battaglini and Harstad 2016) *At any time t , the utility of country $i \in N$ is independent of all past stocks and can be represented by the continuation value function $v_{i,t} = \sum_{\tau=t} \delta^{\tau-t} \hat{u}_{i,\tau}$, where*

$$\hat{u}_{i,t} \equiv -\frac{b}{2} d_{i,t}^2 - \delta \frac{K}{2} R_{i,t+1}^2 - C \sum_{j \in N} (\bar{y}_i - d_{j,t} - \delta R_{j,t+1}), \quad (6)$$

with

$$d_{i,t} \equiv \bar{y}_i - g_{i,t} - R_{i,t}, \quad K \equiv k (1 - \delta q_R^2) e^{\rho\Lambda}, \quad C \equiv \frac{c}{1 - \delta q_G}.$$

The variable $d_{i,t}$ represents energy reduction (relative to the satiation point \bar{y}_i), K is an aggregated cost parameter and C represents the social cost of carbon.

Throughout the paper we restrict our attention to Markov-perfect equilibria (MPE) in pure strategies. For later use as benchmarks we briefly characterize the first-best allocation and the non-cooperative MPE. The latter is referred to as business as usual (BAU). The first-best allocation follows from maximizing $\sum_{j \in N} v_{j,t}$ from (6) with respect to $d_{i,t}$ and $R_{i,t+1}$ which yields

$$-bd_{i,t} + nC = 0 \quad \Leftrightarrow \quad d_{i,t} = n \frac{C}{b} \quad \forall t \geq 1, \quad (7)$$

$$\delta(-KR_{i,t+1} + nC) = 0 \quad \Leftrightarrow \quad R_{i,t+1} = n\frac{C}{K} \quad \forall t \geq 1. \quad (8)$$

In contrast, in BAU each country i maximizes $v_{i,t}$ from (6) with respect to $d_{i,t}$ and $R_{i,t+1}$. The associated first-order conditions can be rearranged to

$$-bd_{i,t} + C = 0 \quad \Leftrightarrow \quad d_{i,t} = \frac{C}{b} \quad \forall t \geq 1, \quad (9)$$

$$\delta(-KR_{i,t+1} + C) = 0 \quad \Leftrightarrow \quad R_{i,t+1} = \frac{C}{K} \quad \forall t \geq 1. \quad (10)$$

Comparing (7)-(8) and (9)-(10) shows that emissions are inefficiently high and investments are inefficiently low in BAU. All countries suffer a welfare loss in BAU, because non-cooperative governments ignore the negative impact of their emissions and the positive impact of their green energy investments on all other countries. The ratio between energy reduction and green investment both in the first best and in BAU is

$$\frac{d_{i,t}}{R_{i,t}} = x \equiv \frac{K}{b} \quad \forall t \geq 1. \quad (11)$$

The parameter x reflects the marginal costs of investments relative to the energy preference and is referred to as *relative cost of technology*.

3 The dynamic game

In the sequel, we analyze a game between coalition countries and non-signatories when contracts are *incomplete*, i.e. coalition countries commit on emissions but not on investments. Non-signatories choose their emissions and green investments non-cooperatively. Since the stocks do not affect the countries' choices, BAU emissions and BAU investments are dominant strategies when countries stay outside the coalition.

Coalition countries choose non-cooperatively investments due to the incomplete contract. If m countries have agreed to join a climate coalition, each coalition country $i \in M \subseteq N$ makes pledges $z_{i,t}$ to curb emissions below BAU emissions $g_{i,t}^{BAU}$, such that its emissions

commitment is given by

$$g_{i,t} = g_{i,t}^{BAU} - z_{i,t}. \quad (12)$$

It is straightforward to show that from a coalition country's perspective, it is equivalent to choose $z_{i,t}$ or $d_{i,t}$. The coalition countries choose their emissions by maximizing the Nash product, formally

$$d_{i,t}^* = \operatorname{argmax}_{d_{i,t}} \prod_{j \in M} v_{j,t}(d_{i,t}, \mathbf{d}_{-i,t}^*)^{\omega_j^j}, \quad (13)$$

where $\mathbf{d}_{-i,t}^*$ is the vector of other coalition countries' ($j \in M \setminus i$) equilibrium energy reduction. In (13), $\omega_j^j / \omega_i^i \in [0, 1)$ is country j 's bargaining power vis-a-vis country i . Presupposed countries are symmetric, (13) is equivalent to

$$d_{i,t}^* = \operatorname{argmax}_{d_{i,t}} \left[v_{i,t}(d_{i,t}, \mathbf{d}_{-i,t}^*) + \omega \sum_{j \in M \setminus i} v_{j,t}(d_{i,t}, \mathbf{d}_{-i,t}^*) \right]. \quad (14)$$

The terms in square brackets in (14) can be interpreted as welfare function of coalition country i , where $\omega \equiv \omega_j^j / \omega_i^i$ is the relative welfare weight of the other coalition countries ($j \in M \setminus i$). For the polar case $\omega = 1$ coalition countries maximize the symmetric Nash product or tantamount the sum of welfares. We refer to (14) with $\omega < 1$ as asymmetric Nash (bargaining) solution and to (14) with $\omega = 1$ as symmetric Nash (bargaining) solution.

Harstad (2021) has shown that the Paris Agreement can be modeled as a Pledge-and-review bargaining game whose solution is the asymmetric Nash solution. In contrast, the symmetric Nash solution is to date the standard approach in the literature of self-enforcing climate agreements⁹ and maps Kyoto-Protocol. In the remainder of this section, we are interested in how emissions, investments, the contract length and the stable coalition of climate agreements change upon variations of the welfare weight ω .

The timing of the game is illustrated in Figure 1. If there is no coalition, each country $i \in N$ decides whether to join a coalition or to stay outside. Then, coalition countries

⁹See the literature mentioned in the Introduction.

negotiate on emissions pledges. Next, non-signatories non-cooperatively choose emissions and coalition countries pollute as agreed. Finally, non-signatories and coalition countries non-cooperatively choose investments. If an agreement already exists, the participation decision and the negotiations are omitted.

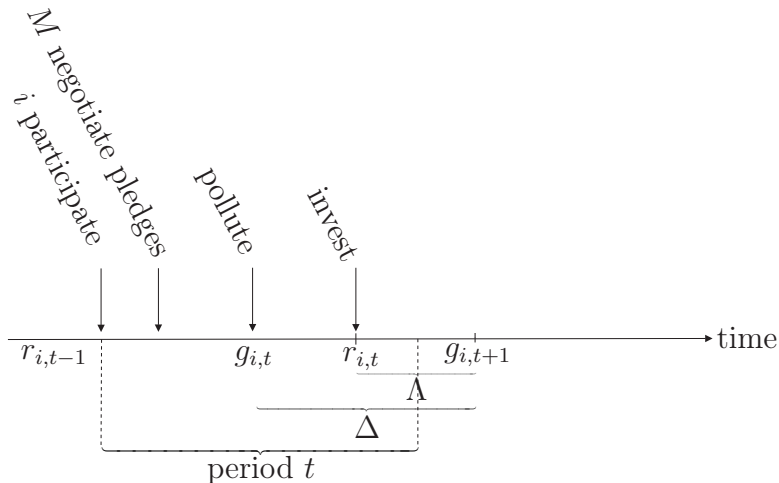


Figure 1: Timing of the game

The MPE of the dynamic game between the coalition and non-signatories is characterized by countries' policies $(d_{i,t}^*(M, T), R_{i,t+1}^*(M, T))_{t=1}^T$, the duration $T^*(M)$ of the agreement and the stable coalition M^* .¹⁰ By virtue of backward induction we first derive the equilibrium emissions and investments for given M and T , then the equilibrium duration for given M and finally the stable coalition.

As mentioned before non-signatories set BAU emissions and BAU investments according to (9) and (10). In the Appendix we derive the coalition countries' emissions and investments of incomplete contracts

$$d_{i,t} = [1 + (m - 1)\omega] \frac{C}{b} \quad \forall i \in M, t \in \{1, \dots, T\}, \quad (15)$$

$$R_{i,t+1} = [1 + (m - 1)\omega] \frac{C}{K}, \text{ but } R_{i,T+1} = \frac{C}{K} \quad \forall i \in M, t \in \{1, \dots, T\}, \quad (16)$$

¹⁰The concept of self-enforcement or stability was originally introduced by D'Aspremont et al. (1983) in the context of cartel formation.

which yields the pledges

$$z_{i,t} = (m - 1)\omega \left(\frac{C}{b} + \frac{C}{K} \right), \text{ but } z_{i,1} = (m - 1)\omega \frac{C}{b} \quad \forall i \in M, t \in \{1, \dots, T\}. \quad (17)$$

In view of (15) and (17) each coalition country reduces its energy consumption relative to its BAU level, and emissions pledges are the higher the larger the welfare weight ω and the larger the coalition m . The larger ω the stronger is the internalization of climate externalities within the coalition, and the larger m the more coalition countries' climate externalities are (partly) internalized. A coalition country's technology investment is greater than its BAU level and increasing in ω and m except for the last period of the agreement.¹¹ In the last period each coalition country realizes that technology investment will be sunk in the next period and chooses BAU investments as non-signatories do. This phenomenon is known as hold-up problem. Recalling that non-signatories choose BAU emissions and BAU investments elucidates the free-riding problem. Coalition countries reduce emissions and step up their green investments to mitigate the climate damage, whereas non-signatories stay at their BAU levels and benefit at zero costs. The larger ω and m the more pronounced are the non-signatories' free-riding incentives.

Next, we determine the optimal duration of the agreement for given coalition M . The coalition country's investments and emissions pledges $(R_{i,t+1}^*, z_{i,t+1}^*)$ from (16) and (17) depend on the contract period at which they are made and offered, respectively. The next lemma specifies the optimal contract length for given coalition size m .

Lemma 2 *Let M^* denote an equilibrium coalition size $m^* \equiv |M^*|$. Then, a coalition of size $m = |M|$, satisfying $M \subseteq M^*$ or $M^* \subseteq M$, finds it optimal to contract for $T(m)$ periods,*

¹¹In the first period of the agreement, $R_{i,1}$ is given, such that $z_{i,1} = d_{i,1} - d_{i,1}^{BAU} = (m - 1)\omega \frac{C}{b}$.

where

$$T(m) = \begin{cases} 1 & \text{if } m < \hat{m}(x, \delta, m^*) \\ \{1, \dots, \infty\} & \text{if } m = \hat{m}(x, \delta, m^*) \\ \infty & \text{if } m > \hat{m}(x, \delta, m^*), \end{cases} \quad (18)$$

with

$$\hat{m}(x, \delta, m^*) \equiv 1 + (m^* - 1) \sqrt{\frac{x + \delta}{x + 1}} < m^*.$$

According to Lemma 1 there is a threshold coalition size \hat{m} such that the coalition of size m concludes a short-term agreement ($T = 1$) if $m < \hat{m}$, and a long-term agreement ($T = \infty$) if $m > \hat{m}$. Each country expects that once the current contract expires, the next contract will be concluded by the equilibrium coalition. For given contract duration, each coalition country's utility is increasing in the coalition size, and for a given coalition size, each coalition country's utility is increasing in the contract duration due to the hold-up problem. To put it differently, the underinvestment of the hold-up problem creates a cost of signing a short-term contract, which stands against the benefit of waiting for a larger coalition in the future. A coalition greater than or equal to the equilibrium coalition ($m \geq m^*$) finds it optimal to conclude a long-term agreement ($T = \infty$). However, also a coalition smaller than the equilibrium coalition ($m \in (\hat{m}, m^*)$) can find it optimal to sign a long-term contract if it is not too small: It then forgoes a larger (equilibrium) coalition in the future to prevent the hold-up problem. If the coalition is very small ($m < \hat{m}$), then it signs a short-term contract ($T = 1$) to achieve a larger coalition in the next period. For $m = \hat{m}$, the benefit of a short-term contract, i.e. the larger coalition in the next period, and its cost, i.e. the underinvestment in the current period, exactly cancel out such that any contract duration is an equilibrium. Note that the equilibrium coalition M^* always concludes a long-term agreement ($T(m^*) = \infty$).

It is worth mentioning that $\hat{m}(x, \delta, m^*)$ does not depend on the welfare weight ω .

Variations of ω cause opposing effects: On the one hand, a smaller ω reduces the difference between the investment of each coalition country and that of each non-signatory, which mitigates the hold-up problem. Contracting for just one period and allowing for the equilibrium coalition in the next period becomes more attractive ($\hat{m} \uparrow$). On the other hand, a smaller ω reduces the internalization of climate externalities within any coalition, which reduces the welfare loss of a narrowed coalition. Contracting forever with a coalition smaller than the equilibrium coalition becomes less costly ($\hat{m} \downarrow$). These two effects exactly cancel out.

Finally, we analyze the stability of the climate coalition. When doing so, we have to make a case distinction depending on the contract length in case of deviation. If a single country deviates by not participating, the remaining coalition sets $T = 1$ only if $m^* - 1 \leq \hat{m}(x, \delta, m^*) \Leftrightarrow m^* \leq m_M(x, \delta)$, where

$$m_M(x, \delta) \equiv 1 + \frac{1}{1 - \sqrt{\frac{x+\delta}{x+1}}}. \quad (19)$$

The inequality $m^* \leq m_M(x, \delta)$ is referred to as *discipline constraint*. $m_M(x, \delta)$ depends on the relative cost of technology $x \equiv \frac{K}{b}$. When x increases, the technology investment becomes more expensive and countries rely more on consumption reduction than on technology investment, which mitigates the hold-up problem. Signing a short-term contract if a single country deviates by not participating becomes less expensive and relaxes the participation constraint ($\frac{\partial m_M}{\partial x} > 0$).

If the discipline constraint is violated ($m^* > m_M(x, \delta)$), the coalition signs a long-term contract ($T = \infty$) even if a single country deviates by not participating. In that case the internal and external stability condition, respectively, is given by

$$m^* \leq m_I(\omega) \equiv 1 + 2/\omega \quad \text{and} \quad m^* > 2/\omega. \quad (20)$$

In view of (20), the stable coalition size is determined by $m^* = \lfloor m_I(\omega) \rfloor$, where $\lfloor \cdot \rfloor$ is the function that maps its argument to the largest weakly smaller integer.¹² We refer to $m_I(\omega)$

¹²If the discipline constraint is violated, the incomplete contract is identical to the complete contract in

as *participation constraint I*. The stable coalition size $\lfloor m_I(\omega) \rfloor$ decreases in ω . Recall that for exogenously given coalition M , an increase in ω enhances the internalization of climate externalities within the coalition. As a consequence coalition countries decrease emissions and raise investments. This in turn enhances the free-riding incentives of non-signatories and decreases the size of the stable coalition.

Approximating the coalition size by $m^* = 1 + 2/\omega$ in the MPE of the dynamic game (i.e. for $M = M^*$), the coalition countries' emissions and investments are $d_{i,t}^* = 3\frac{C}{b}$ and $R_{i,t+1}^* = 3\frac{C}{K}$ for $t \in \{1, \dots, \infty\}$, and hence they are independent of ω . Since the stable coalition is the smaller the larger ω , in the MPE total energy consumption and the climate damage are increasing and total technology investment is decreasing in ω , such that each coalition country's utility is decreasing in ω . Note, however, that even if ω is so small that $m^* = n$, total consumption [investment] is $n/3$ times too high [low] compared to the first-best allocation.

If the discipline constraint is satisfied ($m^* \leq m_M(x, \delta)$), then the coalition signs a short-term contract ($T = 1$) if a single country deviates by not participating. In this one-period agreement, the hold-up problem leads to underinvestment such that the punishment of free riding is higher than with $T = \infty$. The hold-up problem serves as credible threat to restrain countries from leaving the coalition. Now, the internal stability condition is given by

$$m^* \leq m_{\bar{I}}(x, \delta, \omega) \equiv 1 + \frac{2/\omega}{1 - \frac{2-\omega}{\omega} \frac{\delta}{x}} > m_I(\omega), \quad (21)$$

whereas the external stability condition remains unchanged $m^* > 2/\omega$. $m_{\bar{I}}(x, \delta, \omega)$ is referred to as *participation constraint \bar{I}* . Presupposed $m_{\bar{I}}(x, \delta, \omega) < \min\{m_M(x, \delta), n\}$ the stable coalition is given by $m^* = \lfloor m_{\bar{I}}(x, \delta, \omega) \rfloor$. Due to the hold-up problem, the participation constraint \bar{I} is less restrictive than the participation constraint I and leads to larger stable coalitions ($m_{\bar{I}} > m_I$). The comparative static effects of increasing ω are qualitatively

terms of emissions, investments, contract length and stable coalition size.

the same in case of the participation constraints \underline{I} and \bar{I} . With the same interpretation as before, increases in ω enhance free-riding incentives and reduce the size of the stable coalition ($\frac{\partial m_{\bar{I}}}{\partial \omega} < 0$). The participation constraint \bar{I} also depends on the relative cost of technology. When x increases, the technology investment becomes more expensive, countries rely more on energy reduction than on green investment and the hold-up problem is mitigated. Thus, deviating by not participating if the remaining coalition signs a one-period contract becomes less costly, which reduces the size of the stable coalition ($\frac{\partial m_{\bar{I}}}{\partial x} < 0$). A complete characterization of the stable coalition is provided in¹³

Proposition 1 M^* is an equilibrium coalition if and only if either $m^* = \lfloor m_{\underline{I}}(\omega) \rfloor$ or $\lfloor m_{\underline{I}}(\omega) \rfloor < m^* \leq \min\{n, m(x, \delta, \omega)\}$, where

$$m(x, \delta, \omega) = \min\{m_M(x, \delta), m_{\bar{I}}(x, \delta, \omega)\} = \begin{cases} m_M(x, \delta) & \text{if } x < \hat{x}(\delta, \omega) \\ m_{\bar{I}}(x, \delta, \omega) & \text{if } x \geq \hat{x}(\delta, \omega), \end{cases} \quad (22)$$

with

$$\hat{x}(\delta, \omega) \equiv \frac{1 + \delta + \sqrt{(1 + \delta)^2 + \frac{4\delta\omega(4-\omega)}{(2-\omega)^2}}}{\frac{2\omega(4-\omega)}{(2-\omega)^2}} \in \left[\max\left\{\frac{1}{3}, \frac{2-\omega}{\omega}\delta\right\}, \frac{2-\omega}{\omega} \right], \quad \frac{\partial \hat{x}}{\partial \delta} > 0, \quad \frac{\partial \hat{x}}{\partial \omega} < 0.$$

Proposition 1 shows that either the participation constraint \underline{I} , the participation constraint \bar{I} or the discipline constraint $m_M(x, \delta)$ is relevant for the stable coalition. One constraint of the set $\{m_M(x, \delta), m_{\underline{I}}(\omega), m_{\bar{I}}(x, \delta, \omega)\}$ ‘binds’ and determines the size of the stable coalition.

4 Kyoto Protocol versus Paris Agreement

In this section, we compare the Kyoto Protocol with the Paris Agreement in our dynamic game. We assume that the total number of countries is $n = 197$. The symmetric Nash solution ($\omega = 1$) reflects the Kyoto-Protocol, whereas the asymmetric Nash solution with $\omega < 0.5$ reflects the Paris Agreement. Since the Kyoto Protocol has been signed by 37

¹³Proposition 1 is proved in the Appendix.

countries (Canada withdrew in 2012) and the Paris Agreement has been implemented by 195 countries,¹⁴ we consider those economies which provide $m^* = 37$ if $\omega = 1$ and $m^* = 195$ if $\omega < 0.5$ as stable coalitions in the dynamic game. There are three types of feasible economies, denoted as economies $\mathcal{E}_1 - \mathcal{E}_3$, that are different with respect to the binding constraints. Table 1 provides an overview of the relevant constraints in economies $\mathcal{E}_1 - \mathcal{E}_3$. For each economy we present an example.^{15,16}

Economies	Paris Agreement ($\omega < 0.5$)	Kyoto Protocol ($\omega = 1$)
\mathcal{E}_1	discipline constraint	participation constraint \bar{I}
\mathcal{E}_2	participation constraint \underline{I}	participation constraint \bar{I}
\mathcal{E}_3	participation constraint \underline{I}	discipline constraint

Table 1: Binding constraints in the feasible economies \mathcal{E}_1 - \mathcal{E}_3

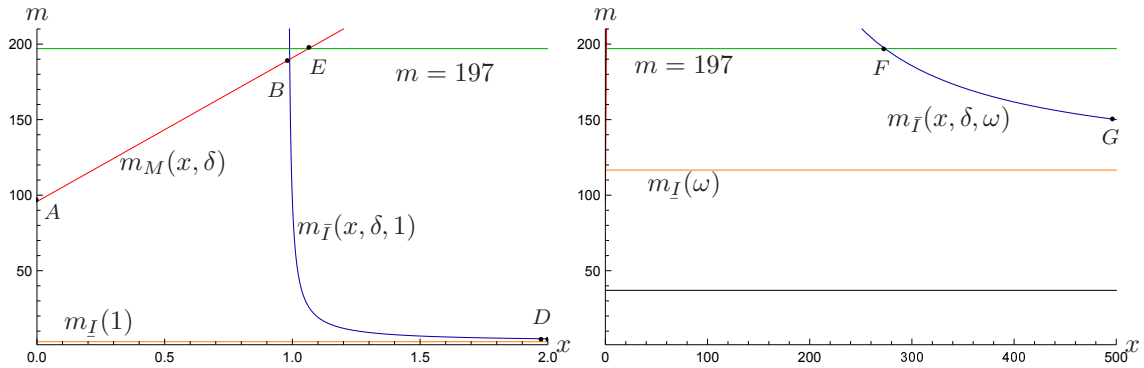


Figure 2: Stable coalitions in Example 1 ($\delta = 0.979$, $\omega = 0.0174$)

We begin with numerical Example 1 which represents the economies \mathcal{E}_1 . The parameter values are $\delta = 0.979$, $\omega = 0.0174$ and $\omega = 1$. Figure 2 illustrates the associated size of the stable coalition m^* in dependence of the relative cost of technology x . The discipline curve m_M is increasing in x and independent of ω , while the participation curve m_I is decreasing in x and it shifts downwards if ω increases. The left panel shows that the stable coalition for $\omega = 1$ is characterized by the polyline ABD . At AB the discipline constraint is binding,

¹⁴Libya has signed and ratified the Paris Agreement, but did not make any pledges.

¹⁵Following Harstad (2020) we assume $\omega < 0.5$ for the Paris Agreement. Therefore, the case that with the Paris Agreement the participation constraint \bar{I} binds, which emerges if and only if $\omega = 0.977$, is excluded.

¹⁶A complete characterization of the economies \mathcal{E}_1 - \mathcal{E}_3 is given in Lemma A1 in the Appendix.

whereas at BD the participation constraint \bar{I} is binding. Reducing ω from 1 to 0.0174 shifts the participation curve $m_{\bar{I}}$ to the right as shown in the right panel of Figure 2. Now the stable coalition lies on the polyline $AEFG$, where AE is in the left panel and FG is in the right panel of Figure 2. For $\omega = 0.0174$ the grand coalition ($m^* = 197$) is stable if x is on the line EF .

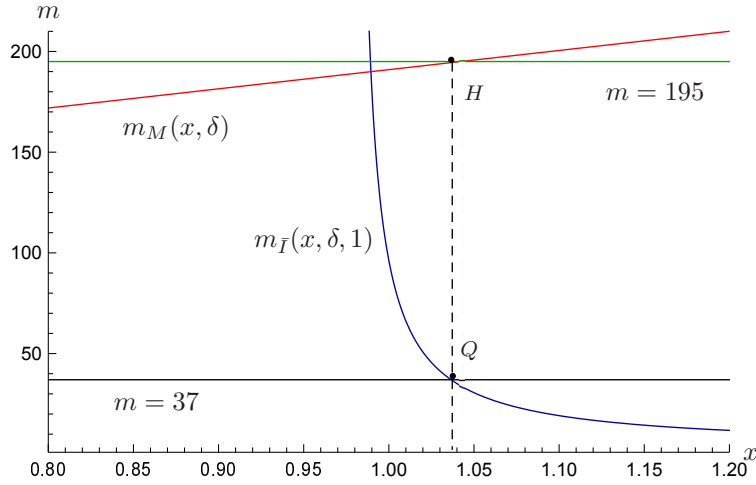


Figure 3: Stable coalitions in Example 1 ($\delta = 0.979$, $\omega = 0.0174$)

Next, consider Figure 3 which is an enlarged segment of Figure 2. In order to compare the Paris Agreement with the Kyoto Protocol we select $x^* = 1.037$ such that the stable coalition is $m^* = 195$ for $\omega = 0.0174$ (point H) and $m^* = 37$ for $\omega = 1$ (point Q). At point Q ($\omega = 1$) the participation constraint \bar{I} is binding, whereas at point H ($\omega = 0.0174$) the discipline constraint is binding (see Table 1). In both agreements the hold-up problem is existent. Closer inspection of the two MPE reveals the following: If the discipline constraint is satisfied as in point H , then the coalition countries' emissions and investments are characterized by $bd_{i,t} = KR_{i,t+1} = [1 + (m_M(x, \delta) - 1)\omega]C = 4.37C$. In contrast, if the participation constraint is satisfied as in point Q , then the coalition countries' emissions and investments are characterized by $bd_{i,t} = KR_{i,t+1} = [1 + \frac{2x\omega}{\omega(x+\delta)-2\delta}]C = 37C$. In the transition from H ($\omega = 0.0174$) to Q ($\omega = 1$) there are two countervailing effects. On the one side each coalition country emits less and invests more. On the other side the coalition size decreases such that more countries free ride. In the transition from the Kyoto Protocol to the

Paris Agreement, the stable coalition becomes broader and shallower. Aggregate emissions increase and aggregate investments decline. In Example 1, ω is chosen such that a coalition country's welfare in the Kyoto Protocol exactly coincides with a coalition country's welfare in the Paris Agreement. Reducing [enhancing] ω below [above] the threshold $\omega = 0.0174$ increases [decreases] a coalition country's emissions and reduces [enhances] its investments such that each coalition country's welfare decreases [increases].^{17,18} In the Appendix, we prove that the economies \mathcal{E}_1 satisfy $x = 1.037$, $\delta = 0.979$ and $\omega \in [\frac{1}{97}, 0.5)$. We summarize these results in

Proposition 2 *In economies \mathcal{E}_1 the welfare of a coalition country is higher with the Kyoto Protocol ($m^* = 37, \omega = 1$) than with the Paris Agreement ($m^* = 195, \omega < 0.5$) if and only if $\omega < 0.0174$.*

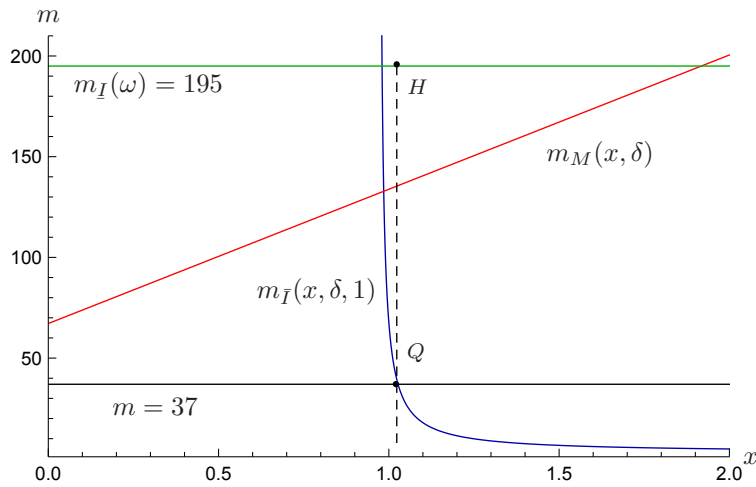


Figure 4: Stable coalitions in Example 2 ($\delta = 0.97$, $\omega = \frac{1}{97}$)

In economies \mathcal{E}_2 for the Kyoto Protocol the participation constraint \bar{I} is still binding but for the Paris Agreement the participation constraint \underline{I} is binding. The hold-up problem

¹⁷Observe that reducing ω below 0.0174 does not change Figure 3. Only curve FG in the right panel of Figure 2 is further shifted to the right.

¹⁸If $\omega = 0.0174$ each non-signatory's welfare is higher with the Kyoto Protocol than with the Paris Agreement. Recall that non-signatories always set BAU emissions and BAU investments. Each non-signatory's welfare depends over aggregate emissions indirectly on the welfare weight ω . Aggregate emissions are decreasing in ω . Thus, there exists a second threshold of ω such that a non-signatory's welfare is higher with the Kyoto Protocol than with the Paris Agreement if and only if $\omega < 0.0352$. See Lemma A2 in the Appendix.

is existent in the Kyoto Protocol, but non-existent in the Paris Agreement. To shift the participation line m_I to the top such that $m = 195$, the welfare weight ω has to be reduced to $\omega = \frac{1}{97}$. Economies \mathcal{E}_2 satisfy $x \in (\frac{35}{37}, \frac{18}{17}]$, $\delta = \frac{17}{18}x$ and $\omega = \frac{1}{97}$. In Example 2, which is illustrated in Figure 4, we have chosen $\delta = 0.97$. Again point H reflects the Paris Agreement and point Q captures the Kyoto Protocol. The small welfare weight $\frac{1}{97} < 0.0174$ implies that a coalition country's welfare decreases in the transition from the Kyoto Protocol to the Paris Agreement.¹⁹

Proposition 3 *In economies \mathcal{E}_2 the welfare of a coalition country is higher with the Kyoto Protocol ($m^* = 37, \omega = 1$) than with the Paris Agreement ($m^* = 195, \omega = \frac{1}{97}$).*

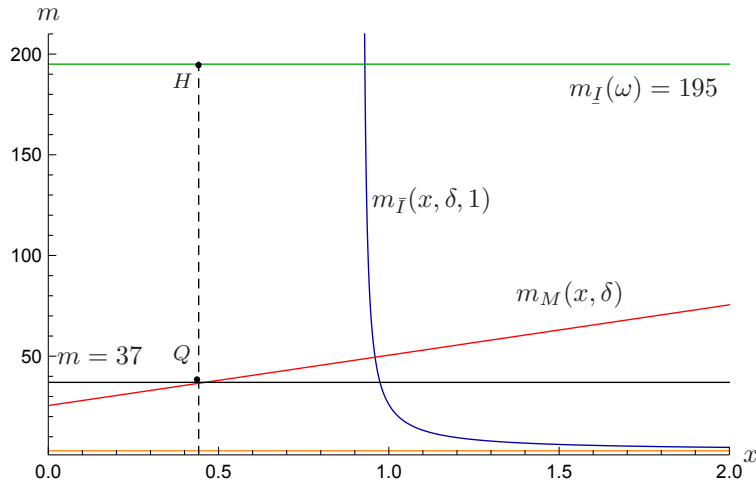


Figure 5: Stable coalitions in Example 3 ($\delta = 0.92, \omega = \frac{1}{97}$)

Finally, we turn to economies \mathcal{E}_3 . In these economies the participation constraint I is binding for the Paris Agreement and the discipline constraint is binding for the Kyoto Protocol. Again, the hold-up problem is existent in the Kyoto Protocol, but non-existent in the Paris Agreement. To ensure that the participation constraint I provides $m^* = 195$, we have to set $\omega = \frac{1}{97}$. Furthermore, the stable coalition $m^* = 37$ lies on the discipline constraint if $x \in (0, \frac{35}{37}]$ and $\delta = \frac{1225}{1296} - \frac{71}{1296}x$ and $\omega = \frac{1}{97}$. Example 3, in which we have set $\delta = 0.92$, is depicted in Figure 5 with H and Q being the MPE for $\omega = \frac{1}{97}$ and $\omega = 1$, respectively. Due to $\omega < 0.0174$ we infer

¹⁹For the proof of that claim we refer to Lemma A2 in the Appendix.

Proposition 4 *In economies \mathcal{E}_3 the welfare of a coalition country is higher with the Kyoto Protocol ($m^* = 37, \omega = 1$) than with the Paris Agreement ($m^* = 195, \omega = \frac{1}{97}$).*

The set of feasible economies is illustrated in Figure 6. The economies \mathcal{E}_2 [\mathcal{E}_3] are on the line AB [BG]. In these economies the Kyoto Protocol performs better for coalition countries (Proposition 3 and 4). The line ED captures the economies \mathcal{E}_1 and the point F is the ‘threshold’ economy \mathcal{E}_1 satisfying $\omega = 0.0174$. For all economies \mathcal{E}_1 on the line FD [EF] a coalition country’s welfare is higher [lower] with the Kyoto Protocol than with the Paris Agreement (Proposition 2). The welfare rankings apply not only for coalition countries but also for non-signatories. To sum up, in the economies on the lines AB , BG and FD the Kyoto Protocol is Pareto superior to the Paris Agreement. Figure 6 shows that this set of economies is considerably larger than the set of economies in which the Kyoto Protocol is Pareto inferior to the Paris Agreement. There exists only one discount factor for which the Paris Agreement performs better and this will only happen if the welfare weight ω is sufficiently large.

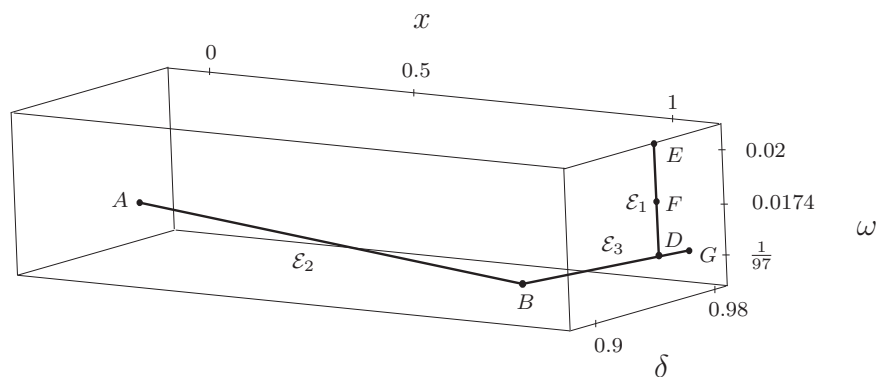


Figure 6: Feasible economies

Finally, it is worth mentioning that our results of Proposition 1-4 are robust with respect to reversing the timing of investments and emissions decisions, and with respect to the time-dependence of pledges. So far, we have assumed that coalition countries set time-dependent pledges. The analysis of time-independent pledges can be found in the Online Appendix.

5 Concluding remarks

The present paper has compared the Kyoto Protocol with the Paris Agreement in a dynamic game in which countries choose emissions reductions and investments in green energy, decide whether to join a climate coalition and negotiate the duration of the agreement. It is shown that the stable coalition is large in the Paris Agreement and small in the Kyoto Protocol, but coalition countries' efforts to reduce emissions are much stronger in the Kyoto Protocol. It turns out that there is a large set of feasible economies in which world emissions are lower and countries' welfare is higher in the Kyoto Protocol than in the Paris Agreement. Our results contrast with Harstad (2020) who finds that countries prefer the Paris Agreement. The conclusion of our analysis is not that countries should go back to the Kyoto Protocol. Instead, our analysis may be interpreted in the way that countries' efforts to mitigate climate change with the Paris Agreement are even smaller than with the Kyoto Protocol and thus still not strong enough to keep global warming below 2° Celsius.

The present analysis can be extended in various directions. Country-specific asymmetries, especially the distinction between industrialized and developing countries, have played an important role in the transition from the Kyoto Protocol to the Paris Agreement. In the same vein, transfers between asymmetric countries may affect the performance of the asymmetric and symmetric Nash solutions in coalition formation games differently. Finally, trade sanctions and positive spillovers from research and development of breakthrough technologies stand on the agenda for future research when comparing the asymmetric with the symmetric Nash solution in the analysis of self-enforcing climate agreements.

References

- Bang, G., Hovi, J. and T. Skodvin (2016): The Paris Agreement: Short-term and long-term effectiveness, *Politics and Governance* 4, 209-218.
- Barrett, S. (2006): Climate treaties and 'breakthrough' technologies, *American Economic*

Review Papers and Proceedings 96, 22-25.

Barrett, S. (1994): Self-enforcing international environmental agreements, *Oxford Economic Papers* 46, 878-894.

Battaglini, M. and B. Harstad (2016): Participation and duration of environmental agreements, *Journal of Political Economy* 124(1), 160-204.

Bayramoglu, B., Finus, M. and J.-F. Jacques (2018): Climate agreements in a mitigation adaptation game, *Journal of Public Economics* 165, 101-113.

Caparrós, A. (2020): Pledge and implement bargaining in the Paris Agreement on climate change, online at: http://ipp.csic.es/sites/default/files/content/workpaper/2020/ipp_wp_3_2020_pledge.pdf.

Carraro, C. and D. Siniscalco (1993): Strategies for the international protection of the environment, *Journal of Public Economics* 52, 309-328.

D'Aspremont, C., Jacquemin, A., Gabszewicz, J.J. and J.A. Weymark (1983): On the stability of collusive price leadership, *Canadian Journal of Economics* 16, 17-25.

Dutta, P.K. and R. Radner (2004): Self-enforcing climate-change treaties, *Proceedings of the National Academy of Sciences* 101, 4746-4751.

Harstad, B. (2020): Pledge-and-review bargaining: From Kyoto to Paris, online at: <http://www.sv.uio.no/econ/personer/vit/bardh/dokumenter/ndc.pdf>, *Economic Journal*, revise and resubmit.

Harstad, B. (2021): A theory of pledge-and-review bargaining, online at: <https://www.sv.uio.no/econ/personer/vit/bardh/dokumenter/prb.pdf>, *Journal of Economic Theory*, revise and resubmit.

Hoel, M. (1992): International environmental conventions: the case of uniform reductions of emissions, *Environmental and Resource Economics* 2, 141-159.

- Hoel, M. and Aart de Zeeuw (2010): Can a focus on breakthrough technologies improve the performance of international environmental agreements?, *Environmental and Resource Economics* 47, 395-406.
- Hulme, M. (2016): The climate research agenda after Paris: Should 1.5 degrees change anything?, *Nature* 453(19), 268-269.
- Karp, L. and H. Sakamoto (2021): Sober optimism and the formation of international environmental agreements, *Journal of Economic Theory* 197, 105321.
- Kornek, U. and O. Edenhofer (2020): The strategic dimension of financing global public goods, *European Economic Review* 127, 103423.
- Kováč, E. and R.C. Schmidt (2021): A simple dynamic climate cooperation model, *Journal of Public Economics* 194, 104329.
- Lippert, S. and J. Tremewan (2021): Pledge-and-review in the laboratory, *Games and Economic Behavior* 130, 179-195.
- McEvoy, D.M. and M. McGinty (2018): Negotiating a uniform emissions tax in international environmental agreements, *Journal of Environmental Economics and Management* 90, 217-231.
- Rogelj, J., den Elzen, M., Höhne, N., Fransen, T., Fekete, H., Winkler, H., Schaeffer, R., Sha, F., Riahi, K. and M. Meinshausen (2016): Paris Agreement climate proposals need a boost to keep warming well below 2° C, *Nature* 534, 631-639.
- Rubio, S.J. and B. Casino (2005): Self-enforcing international environmental agreements with a stock pollutant, *Spanish Economic Review* 7, 89-109.
- Rubio, S.J. and A. Ulph (2006): Self-enforcing agreements and international trade in greenhouse emission rights, *Oxford Economic Papers* 58, 233-263.
- Rubio, S.J. and A. Ulph (2007): An infinite-horizon model of dynamic membership of international environmental agreements, *Journal of Environmental Economics and Man-*

Appendix

Derivation of equations (15)-(17)

Rewriting (6) yields

$$\begin{aligned}
 v_i &= \sum_{t=1}^T \delta^{t-1} \left[-\frac{b}{2} d_{i,t}^2 - \delta \frac{K}{2} R_{i,t+1}^2 - C \sum_{j \in N} (\bar{y}_i - d_{j,t} - \delta R_{j,t+1}) \right] + \delta^T v_i \\
 &= \sum_{t=1}^T \delta^{t-1} \left[-\frac{b}{2} (\bar{y}_i - g_{i,t} - R_{i,t})^2 - \delta \frac{K}{2} R_{i,t+1}^2 - C \sum_{j \in N} (g_{j,t} + R_{j,t} - \delta R_{j,t+1}) \right] + \delta^T v_i \\
 &= \sum_{t=2}^T \delta^{t-1} \left[-\frac{b}{2} (\bar{y}_i - g_{i,t} - R_{i,t})^2 - \frac{K}{2} R_{i,t}^2 - C \sum_{j \in N} g_{j,t} \right] + \delta^T v_i \\
 &\quad - \frac{b}{2} (\bar{y}_i - g_{i,1} - R_{i,1})^2 - \delta^T \frac{K}{2} R_{i,T+1}^2 - C \sum_{j \in N} (g_{j,1} + R_{j,1} - \delta^T R_{j,T+1}). \tag{A1}
 \end{aligned}$$

The participants' technology investment is given by maximizing v_i from (A1) for given $g_{i,t}$ over $R_{i,t}$:

$$\delta^{t-1} \left[b (\bar{y}_i - g_{i,t} - R_{i,t}) - K R_{i,t} \right] = 0 \quad \Leftrightarrow \quad R_{i,t} = \frac{b}{b+K} (\bar{y}_i - g_{i,t}) \quad \forall t \in \{2, \dots, T\}, \tag{A2}$$

$$-\delta^T K R_{i,T+1} + \delta^T C = 0 \quad \Leftrightarrow \quad R_{i,T+1} = \frac{C}{K}. \tag{A3}$$

Substituting (A2) and (A3) into (A1), we get

$$\begin{aligned}
 v_i &= \sum_{t=2}^T \delta^{t-1} \left[-\frac{b}{2} \left(\frac{K}{b+K} (\bar{y}_i - g_{i,t}) \right)^2 - \frac{K}{2} \left(\frac{b}{b+K} (\bar{y}_i - g_{i,t}) \right)^2 - C \sum_{j \in N} g_{j,t} \right] + \delta^T v_i \\
 &\quad - \frac{b}{2} (\bar{y}_i - g_{i,1} - R_{i,1})^2 - \delta^T \frac{K}{2} \left(\frac{C}{K} \right)^2 - C \sum_{j \in N} \left(g_{j,1} + R_{j,1} - \delta^T \frac{C}{K} \right). \tag{A4}
 \end{aligned}$$

The participants' fossil fuel consumption is given by maximizing $v_i + \omega \sum_{j \in M \setminus i} v_j$ from (A4) over $g_{i,t}$:

$$b (\bar{y}_i - g_{i,1} - R_{i,1}) - \Omega(m)C = 0 \quad \Leftrightarrow \quad \bar{y}_i - g_{i,1} - R_{i,1} = \Omega(m) \frac{C}{b}, \tag{A5}$$

$$\begin{aligned} & \delta^{t-1} \left[b \left(\frac{K}{b+K} \right)^2 (\bar{y}_i - g_{i,t}) + K \left(\frac{b}{b+K} \right)^2 (\bar{y}_i - g_{i,t}) - \Omega(m)C \right] = 0 \\ \Leftrightarrow & \bar{y}_i - g_{i,t} = \Omega(m) \frac{C(b+K)}{bK} \quad \forall t \in \{2, \dots, T\}, \end{aligned} \quad (\text{A6})$$

where $\Omega(m) \equiv 1 + (m-1)\omega$. Substituting (A5) and (A6) into (A2) and $d_{i,t} = \bar{y}_i - g_{i,t} - R_{i,t}$, we get

$$R_{i,t} = \frac{b}{b+K} \Omega(m) \frac{C(b+K)}{bK} = \Omega(m) \frac{C}{K} \quad \forall t \in \{2, \dots, T\}, \quad (\text{A7})$$

$$d_{i,t} = \bar{y}_i - g_{i,t} - R_{i,t} = \Omega(m) \frac{C}{b} \quad \forall t \in \{1, \dots, T\}. \quad (\text{A8})$$

Equation (15) then follows from (A8), and equation (16) then follows from (A3) and (A7). From (12), $d_{i,t} = \bar{y}_i - g_{i,t} - R_{i,t}$, (9) and (10) for BAU emissions and investments, and (A3), (A7) and (A8) for the coalition countries' emissions and investments, we get

$$\begin{aligned} z_{i,t} &= g_{i,t}^{BAU} - g_{i,t} = [\bar{y}_i - d_{i,t}^{BAU} - R_{i,t}^{BAU}] - [\bar{y}_i - d_{i,t} - R_{i,t}] \\ &= \begin{cases} [\bar{y}_i - \frac{C}{b} - R_{i,1}] - [\bar{y}_i - \Omega(m) \frac{C}{b} - R_{i,1}] & \text{if } t = 1 \\ [\bar{y}_i - \frac{C}{b} - \frac{C}{K}] - [\bar{y}_i - \Omega(m) \frac{C}{b} - \Omega(m) \frac{C}{K}] & \text{if } t \in \{2, \dots, T\}. \end{cases} \end{aligned} \quad (\text{A9})$$

Equation (17) then follows from (A9).

Proof of Lemma 2

We first prove that the optimal contract duration T^* of the equilibrium coalition M^* is infinity. If m^* countries contract for $T^* \leq \infty$ periods, each participant's continuation value from (6), (9) and (10) for $i \notin M$, and (15) and (16) for $i \in M$ is given by

$$\begin{aligned} v(m^*, T^*) &= \sum_{t=1}^{T^*-1} \delta^{t-1} \left\{ -\frac{b}{2} \left(\Omega(m^*) \frac{C}{b} \right)^2 - \delta \frac{K}{2} \left(\Omega(m^*) \frac{C}{K} \right)^2 \right. \\ &\quad \left. - C \left[\bar{y} - (m^* \Omega(m^*) + n - m^*) \left(\frac{C}{b} + \frac{\delta C}{K} \right) \right] \right\} \\ &\quad + \delta^{T^*-1} \left\{ -\frac{b}{2} \left(\Omega(m^*) \frac{C}{b} \right)^2 - \delta \frac{K}{2} \left(\frac{C}{K} \right)^2 \right. \\ &\quad \left. - C \left[\bar{y} - (m^* \Omega(m^*) + n - m^*) \frac{C}{b} - n \frac{\delta C}{K} \right] \right\} + \delta^{T^*} v(m^*, T^*) \end{aligned}$$

$$\begin{aligned}
&= -\frac{1-\delta^{T^*}}{1-\delta}C \left[\bar{y} - C \left(m^*\Omega(m^*) - \frac{\Omega(m^*)^2}{2} + n - m^* \right) \left(\frac{1}{b} + \frac{\delta}{K} \right) \right] \\
&\quad - \delta^{T^*} \frac{C^2}{2K} [\Omega(m^*) - 1][2m^* - 1 - \Omega(m^*)] + \delta^{T^*} v(m^*, T^*) \\
&= -\frac{C}{1-\delta} \left[\bar{y} - C \left(m^*\Omega(m^*) - \frac{\Omega(m^*)^2}{2} + n - m^* \right) \left(\frac{1}{b} + \frac{\delta}{K} \right) \right] \\
&\quad - \frac{\delta^{T^*}}{1-\delta^{T^*}} \frac{C^2}{2K} [\Omega(m^*) - 1][2m^* - \Omega(m^*) - 1], \tag{A10}
\end{aligned}$$

where $\bar{y} \equiv \sum_{j \in N} \bar{y}_j$. From (A10), we get the difference in each participant's continuation value between $T^* = \infty$ and $T^* < \infty$

$$v(m^*, T^* = \infty) - v(m^*, T^* < \infty) = \frac{\delta^{T^*}}{1-\delta^{T^*}} \frac{C^2}{2K} [\Omega(m^*) - 1][2m^* - \Omega(m^*) - 1] > 0, \tag{A11}$$

such that $T^* < \infty$ cannot be optimal.

Now we derive the optimal contract duration T of a given coalition M . If m countries contract for T periods and $T^* = \infty$, each participant's continuation value from (6), (9) and (10) for $i \notin M$, and (15) and (16) for $i \in M$ is given by

$$\begin{aligned}
v(m, T) &= -\frac{1-\delta^T}{1-\delta}C \left[\bar{y}_i - C \left(m\Omega(m) - \frac{\Omega(m)^2}{2} + n - m \right) \left(\frac{1}{b} + \frac{\delta}{K} \right) \right] \\
&\quad - \delta^T \frac{C^2}{2K} [\Omega(m) - 1][2m - \Omega(m) - 1] \\
&\quad - \frac{\delta^T}{1-\delta}C \left[\bar{y}_i - C \left(m^*\Omega(m^*) - \frac{\Omega(m^*)^2}{2} + n - m^* \right) \left(\frac{1}{b} + \frac{\delta}{K} \right) \right]. \tag{A12}
\end{aligned}$$

Note that the derivative of $v(m, T)$ with respect to T or, equivalently, with respect to $-\delta^T$ is always negative if and only if

$$\begin{aligned}
&\frac{C^2}{1-\delta} \left(m\Omega(m) - \frac{\Omega(m)^2}{2} + n - m \right) \left(\frac{1}{b} + \frac{\delta}{K} \right) + \frac{C^2}{2K} [\Omega(m) - 1][2m - \Omega(m) - 1] \leq \\
&\frac{C^2}{1-\delta} \left(m^*\Omega(m^*) - \frac{\Omega(m^*)^2}{2} + n - m^* \right) \left(\frac{1}{b} + \frac{\delta}{K} \right) \\
\Leftrightarrow &\frac{x+1}{x+\delta} \frac{(m-1)^2\omega(2-\omega)}{2} - \frac{1}{2} \leq \frac{(m^*-1)^2\omega(2-\omega)}{2} - \frac{1}{2}, \tag{A13}
\end{aligned}$$

where $x \equiv K/b$. From (A13), the optimal contract duration is one period if $m < \hat{m}(x, \delta, m^*)$, infinity if $m > \hat{m}(x, \delta, m^*)$, and arbitrary if $m = \hat{m}(x, \delta, m^*)$, where $\hat{m}(x, \delta, m^*)$ is defined in Lemma 2. QED

Proof of Proposition 1

We first derive the condition for external stability. If a non-participant joins in equilibrium, then $m = m^* + 1$, which is not beneficial to him if his continuation value in case of participation for $m = m^* + 1$ from (A10) for $T^* = \infty$ falls short of his continuation value in case of non-participation for $m = m^*$ from (6) for $d_{i,t}$ and $R_{i,t+1}$ from (9) and (10):

$$\begin{aligned}
& -\frac{C}{1-\delta} \left[\bar{y} - C \left((m^* + 1)\Omega(m^* + 1) - \frac{\Omega(m^* + 1)^2}{2} + n - (m^* + 1) \right) \left(\frac{1}{b} + \frac{\delta}{K} \right) \right] < \\
& -\frac{C}{1-\delta} \left[\bar{y} - C \left(m^*\Omega(m^*) - \frac{1}{2} + n - m^* \right) \left(\frac{1}{b} + \frac{\delta}{K} \right) \right] \\
\Leftrightarrow & \left[(m^* + 1)\Omega(m^* + 1) - \frac{\Omega(m^* + 1)^2}{2} - m^*\Omega(m^*) - \frac{1}{2} \right] \frac{C^2}{1-\delta} \left(\frac{1}{b} + \frac{\delta}{K} \right) < 0 \\
\Leftrightarrow & -m^*\omega^2 \left(m^* - \frac{2}{\omega} \right) \frac{C^2}{2(1-\delta)} \left(\frac{1}{b} + \frac{\delta}{K} \right) < 0, \tag{A14}
\end{aligned}$$

requiring $m^* > 2/\omega$ for external stability.

Now we derive the conditions for internal stability. Suppose $m^* > m_M(x, \delta)$. If a participant deviates in equilibrium, then $m = m^* - 1 > \hat{m}(x, \delta, m^*)$; so $T = \infty$ by Lemma 2. Such a permanent deviation is not beneficial to him if his continuation value in case of participation for $m = m^*$ from (A10) for $T^* = \infty$ exceeds his continuation value in case of non-participation for $m = m^* - 1$ from (6) for $d_{i,t}$ and $R_{i,t+1}$ from (9) and (10):

$$\begin{aligned}
& -\frac{C}{1-\delta} \left[\bar{y} - C \left(m^*\Omega(m^*) - \frac{\Omega(m^*)^2}{2} + n - m^* \right) \left(\frac{1}{b} + \frac{\delta}{K} \right) \right] \geq \\
& -\frac{C}{1-\delta} \left[\bar{y} - C \left((m^* - 1)\Omega(m^* - 1) - \frac{1}{2} + n - (m^* - 1) \right) \left(\frac{1}{b} + \frac{\delta}{K} \right) \right] \\
\Leftrightarrow & \left[m^*\Omega(m^*) - \frac{\Omega(m^*)^2}{2} - (m^* - 1)\Omega(m^* - 1) - \frac{1}{2} \right] \frac{C^2}{1-\delta} \left(\frac{1}{b} + \frac{\delta}{K} \right) \geq 0
\end{aligned}$$

$$\Leftrightarrow (m^* - 1)\omega^2 \left(\frac{2}{\omega} - (m^* - 1) \right) \frac{C^2}{2(1 - \delta)} \left(\frac{1}{b} + \frac{\delta}{K} \right) \geq 0, \quad (\text{A15})$$

requiring $m^* \leq m_I(x, \omega) = 1 + 2/\omega$ for internal stability. From (A14), the coalition is externally stable if $m^* > 2/\omega$, which is only fulfilled for the largest internally stable coalition. Thus, $m^* > m_M(x, \delta)$ implies $m^* = \lceil 1 + 2/\omega \rceil$.

Now suppose $m^* \leq m_M(x, \delta)$. If a participant deviates in equilibrium, then $m = m^* - 1 \leq \hat{m}(x, \delta, m^*)$; so $T = 1$ by Lemma 2, and the participant is expected to join the coalition next period. Such a one-period deviation is not beneficial to him if his one-period utility in case of participation for $m = m^*$ from (A10) for $T^* = \infty$ exceeds his one-period utility in case of non-participation for $m = m^* - 1$ from (6) for $d_{i,t}$ and $R_{i,t+1}$ from (9) and (10):

$$\begin{aligned} & -C \left[\bar{y} - C \left(m^* \Omega(m^*) - \frac{\Omega(m^*)^2}{2} + n - m^* \right) \left(\frac{1}{b} + \frac{\delta}{K} \right) \right] \geq \\ & -\frac{b}{2} \left(\frac{C}{b} \right)^2 - \delta \frac{K}{2} \left(\frac{C}{K} \right)^2 \\ & -C \left[\bar{y} - ((m^* - 1)\Omega(m^* - 1) + n - (m^* - 1)) \frac{C}{b} - n \frac{\delta C}{K} \right] \\ \Leftrightarrow & \left[m^* \Omega(m^*) - \frac{\Omega(m^*)^2}{2} - (m^* - 1)\Omega(m^* - 1) - \frac{1}{2} \right] \frac{C^2}{b} \\ & + \left[m^* \Omega(m^*) - \frac{\Omega(m^*)^2}{2} - m^* + \frac{1}{2} \right] \frac{\delta C^2}{x b} \geq 0 \\ \Leftrightarrow & (m^* - 1)\omega^2 \left[\frac{2}{\omega} + (m^* - 1) \left(\frac{2 - \omega \delta}{\omega x} - 1 \right) \right] \frac{C^2}{2b} \geq 0, \quad (\text{A16}) \end{aligned}$$

requiring $m^* \leq m_{\bar{I}}(x, \delta, \omega) = 1 + \frac{2/\omega}{1 - \frac{2-\omega}{\omega} \frac{\delta}{x}}$ for internal stability. For $x \leq \frac{2-\omega}{\omega} \delta$, any coalition is internally stable. From (A14), the coalition is externally stable if $m^* > 2/\omega$, which is definitely fulfilled for the largest internally stable coalition. Furthermore, $m^* = m_M(x, \delta) < \lceil 1 + 2/\omega \rceil$ would imply $m^* = m_M(x, \delta) \leq 2/\omega$, such that the coalition would not be externally stable and $m^* > m_M(x, \delta)$ would hold. Thus, $m^* \leq m_M(x, \delta)$ implies $m^* \in \left[\lceil 1 + 2/\omega \rceil, \min\{n, m_{\bar{I}}(x, \delta, \omega)\} \right]$.

Comparing $m_M(x)$ from (19) with $m_{\bar{I}}(x, \delta, \omega)$ from (21) yields $m_M(x, \delta) \gtrless m_{\bar{I}}(x, \delta, \omega) \Leftrightarrow x \gtrless \hat{x}(\delta, \omega)$ as defined in Proposition 1, which proves the cases in (22). Differentiating $\hat{x}(\delta, \omega)$ from Proposition 1 with respect to δ and ω , we get

$$\frac{\partial \hat{x}}{\partial \delta} = \frac{1 + \delta + \frac{2\omega(4-\omega)}{(2-\omega)^2} + \sqrt{(1 + \delta)^2 + \frac{4\delta\omega(4-\omega)}{(2-\omega)^2}}}{\frac{2\omega(4-\omega)}{(2-\omega)^2} \sqrt{(1 + \delta)^2 + \frac{4\delta\omega(4-\omega)}{(2-\omega)^2}}} > 0, \quad (\text{A17})$$

$$\frac{\partial \hat{x}}{\partial \omega} = \frac{(1 + \delta)^2 + \frac{2\delta\omega(4-\omega)}{(2-\omega)^2} + (1 + \delta) \sqrt{(1 + \delta)^2 + \frac{4\delta\omega(4-\omega)}{(2-\omega)^2}}}{\frac{\omega^2(4-\omega)^2}{4(2-\omega)} \sqrt{(1 + \delta)^2 + \frac{4\delta\omega(4-\omega)}{(2-\omega)^2}}} < 0, \quad (\text{A18})$$

which proves the signs of the derivatives in Proposition 1. Finally,

$$\hat{x}(\delta, \omega) - \frac{2 - \omega}{\omega} \delta = \frac{4\delta(1 - \delta)/\omega}{\frac{\omega(1+\delta) - 2 + 6\delta}{2 - \omega} + \sqrt{(1 + \delta)^2 + \frac{4\delta\omega(4-\omega)}{(2-\omega)^2}}} \geq 0, \quad (\text{A19})$$

$\hat{x}(0, \omega) = 1/3$ and $\hat{x}(1, \omega) = (2 - \omega)/\omega$ proves the bounds of $\hat{x}(\delta, \omega)$ in Proposition 1. QED

To prove Propositions 2-4, we first characterize the feasible economies \mathcal{E}_1 - \mathcal{E}_3 in Lemma A1 and then derive the welfare difference of each participant and that of each non-participant between the Kyoto Protocol and the Paris Agreement in Lemma A2.

Lemma A1 *Suppose $m^* = 37$ holds for $\omega = 1$ (Kyoto Protocol) and $m^* = 195$ holds for $\omega \leq 0.5$ (Paris Agreement). Then, Table A1 characterizes the feasible economies.*

Economy	Paris Agreement	Kyoto Protocol	$x \in$	$\delta \in$	$\omega \in$
\mathcal{E}_1	$m^* = m_M(x, \delta)$	$m^* = m_{\bar{I}}(x, \delta, 1)$	1.037	0.979	[0.01, 0.5)
\mathcal{E}_2	$m^* = m_{\bar{I}}(\omega)$	$m^* = m_{\bar{I}}(x, \delta, 1)$	(0.946, 1.059]	(0.893, 1]	0.01
\mathcal{E}_3	$m^* = m_{\bar{I}}(\omega)$	$m^* = m_M(x, \delta)$	(0, 0.946]	[0.893, 0.945)	0.01

Table A1: Feasible economies \mathcal{E}_1 - \mathcal{E}_3

Proof of Lemma A1

For the Kyoto Protocol, $m^* = 37$ and $\omega = 1$ imply that either $m^* \leq m_M(x, \delta)$ or $m^* \leq m_{\bar{I}}(x, \delta, \omega = 1)$ is binding. Else, $m^* = m_{\bar{I}}(\omega = 1) = 3$ or $m^* = n = 197$ would hold. First suppose $m^* \leq m_M(x, \delta)$ and $m^* = m_{\bar{I}}(x, \delta, \omega = 1)$ hold for the Kyoto Protocol. From (19)

and (21), we then get

$$m_M(x, \delta) = 1 + \frac{1}{1 - \sqrt{\frac{x+\delta}{x+1}}} \geq 37 \quad \Leftrightarrow \quad \delta \geq \frac{1225 - 71x}{1296}, \quad (\text{A20})$$

$$m_{\bar{I}}(x, \delta, \omega = 1) = 1 + \frac{2x}{x - \delta} = 37 \quad \Leftrightarrow \quad x = \frac{18}{17}\delta. \quad (\text{A21})$$

Substituting (A21) into (A20) and rearranging yields $\delta \geq \frac{595}{666} \approx 0.893$, and substituting this into (A21) yields $x \geq \frac{35}{37} \approx 0.946$. $\delta \leq 1$ then implies $\delta \in [\frac{595}{666}, 1]$ and $x \in [\frac{35}{37}, \frac{18}{17}]$.

For the Paris Agreement, $m^* = 195$ and $\omega \leq 0.5$ imply that $m^* \leq m_{\bar{I}}(\omega)$, $m^* \leq m_M(x, \delta)$ or $m^* \leq m_{\bar{I}}(x, \delta, \omega)$ is binding. Else, $m^* = n = 197$ would hold. Substituting $x = \frac{18}{17}\delta$ from (A21) into (19) and (21), we get

$$m_M\left(x = \frac{18}{17}\delta, \delta\right) = 1 + \frac{1}{1 - \sqrt{\frac{35\delta}{18\delta+17}}} = 195 \quad \Leftrightarrow \quad \delta = \frac{633233}{646778} \approx 0.979, \quad (\text{A22})$$

$$m_{\bar{I}}\left(x = \frac{18}{17}\delta, \delta, \omega\right) = 1 + \frac{36}{35\omega - 34} = 195 \quad \Leftrightarrow \quad \omega = \frac{3316}{3395} \approx 0.977. \quad (\text{A23})$$

$m_{\bar{I}}(x = \frac{18}{17}\delta, \delta, \omega) = 195$ cannot hold since $\omega < 0.5$, and $m_M(x = \frac{18}{17}\delta, \delta) = 195$ holds for $\delta \approx 0.979$ and $x = \frac{18}{17}\delta \approx 1.037$. Finally, $m_{\bar{I}}(\omega) = 1 + \frac{2}{\omega} = 195$ holds for $\omega = \frac{1}{97} \approx 0.010$. Thus, economies \mathcal{E}_1 are characterized by $m_{\bar{I}}(x, \delta, \omega = 1) = 37$, $m_M(x, \delta) = 195$ and $m_{\bar{I}}(\omega) \geq 195$, which implies the values in the second line of Table A1, and economies \mathcal{E}_2 are characterized by $m_{\bar{I}}(x, \delta, \omega = 1) = 37$, $m_M(x, \delta) \geq 37$ and $m_{\bar{I}}(\omega) = 195$, which implies the values in the third line of Table A1.

Now suppose $m^* = m_M(x, \delta)$ and $m^* \leq m_{\bar{I}}(x, \delta, \omega = 1)$ hold for the Kyoto Protocol. From (19) and (21), we then get

$$m_M(x, \delta) = 1 + \frac{1}{1 - \sqrt{\frac{x+\delta}{x+1}}} = 37 \quad \Leftrightarrow \quad \delta = \frac{1225 - 71x}{1296}, \quad (\text{A24})$$

$$m_{\bar{I}}(x, \delta, \omega = 1) = 1 + \frac{2x}{x - \delta} \geq 37 \quad \Leftrightarrow \quad x \leq \frac{18}{17}\delta. \quad (\text{A25})$$

Substituting (A25) into (A24) and rearranging yields $\delta \geq \frac{595}{666} \approx 0.893$, and substituting this into (A25) yields $x \leq \frac{35}{37} \approx 0.946$. $x \geq 0$ then implies $\delta \in [\frac{595}{666}, \frac{1225}{1296}]$ and $x \in [0, \frac{35}{37}]$.

For the Paris Agreement, $m^* = 195$ and $\omega \leq 0.5$ imply that $m^* \leq m_I(\omega)$ or $m^* \leq m_I(x, \delta, \omega)$ is binding. Else, $m_M(x, \delta) = 37$ or $m^* = n = 197$ would hold. Substituting $x = \frac{1225-1296\delta}{71}$ from (A24) into (21), we get

$$m_I \left(x = \frac{1225 - 1296\delta}{71}, \delta, \omega \right) = 1 + \frac{142\delta - 2450(1 - \delta)}{142\delta - 1225\omega(1 - \delta)} = 195 \Leftrightarrow \omega = \frac{12478\delta + 1225}{118825(1 - \delta)}. \quad (\text{A26})$$

$m_I \left(x = \frac{1225-1296\delta}{71}, \delta, \omega \right) = 195$ cannot hold since $\delta \geq \frac{595}{666}$ implies $\omega \geq \frac{3316}{3395} > 0.5$. Finally, $m_I(\omega) = 1 + \frac{2}{\omega} = 195$ holds for $\omega = \frac{1}{97} \approx 0.010$. Thus, economies \mathcal{E}_3 are characterized by $m_M(x, \delta) = 37$, $m_I(x, \delta, \omega = 1) \geq 37$ and $m_I(\omega) = 195$, which implies the values in the last line of Table A1. QED

Lemma A2 *Suppose $m^* = 37$ holds for $\omega = 1$ (Kyoto Protocol) and $m^* = 195$ holds for $\omega \leq 0.5$ (Paris Agreement). Then, the welfare of each participant [non-participant] is higher with the Kyoto Protocol than with the Paris Agreement if and only if $\omega < 0.0174$ [$\omega < 0.0352$]. Furthermore, the intertemporal climate damage is smaller with the Kyoto Protocol than with the Paris Agreement if and only if $\omega < 0.0352$.*

Proof of Lemma A2

For the Kyoto Protocol, we have $m^* = 37$ and $\omega = 1$, and for the Paris Agreement, we have $m^* = 195$ and $\omega \leq 0.5$. From (A10) for $T^* = \infty$, we get the welfare difference of each participant

$$v(m = 37, \omega = 1) - v(m = 195, \omega < 1) = \frac{C^2}{1 - \delta} [648 - 18818\omega(2 - \omega)] \left(\frac{1}{b} + \frac{\delta}{K} \right), \quad (\text{A27})$$

which is positive [negative] for $\omega < [>]0.0174$, and from the first line's right-hand side of (A14), we get the welfare difference of each non-participant

$$v(m = 37, \omega = 1) - v(m = 195, \omega < 1) = \frac{C^2}{1 - \delta} [1332 - 37830\omega] \left(\frac{1}{b} + \frac{\delta}{K} \right), \quad (\text{A28})$$

which is positive [negative] for $\omega < [>]0.0352$. Since each nonparticipant always chooses the business-as-usual energy consumption and technology investment, its welfare difference

stems from the difference in the intertemporal climate damage, such that $\omega < [>]0.0352$ implies a smaller [greater] intertemporal climate damage with the Kyoto Protocol than with the Paris Agreement. QED

Proof of Proposition 2

From Lemma A1, $\omega \in [0.1, 0.5]$ holds in economies \mathcal{E}_1 , such that the welfare of each participant is higher with the Kyoto Protocol than with the Paris Agreement if and only if $\omega < 0.0174$ in economies \mathcal{E}_1 from Lemma A2. QED

Proof of Proposition 3

From Lemma A1, $\omega = 0.1$ holds in economies \mathcal{E}_2 , such that the welfare of each participant is higher with the Kyoto Protocol than with the Paris Agreement in economies \mathcal{E}_2 from Lemma A2. QED

Proof of Proposition 4

From Lemma A1, $\omega = 0.1$ holds in economies \mathcal{E}_3 , such that the welfare of each participant is higher with the Kyoto Protocol than with the Paris Agreement in economies \mathcal{E}_3 from Lemma A2. QED

Online Appendix: Time-independent pledges

Derivation of the coalition countries' emissions, investments and pledges

The participants' technology investment is given by (A2) and (A3). Substituting (A2) and (A3) into (A1) for $g_{i,t} = g_{i,t}^{BAU} - z_i$, we get

$$\begin{aligned}
 v_i = & \sum_{t=2}^T \delta^{t-1} \left\{ -\frac{b}{2} \left(\frac{K}{b+K} (\bar{y}_i - g_{i,t}^{BAU} + z_i) \right)^2 - \frac{K}{2} \left(\frac{b}{b+K} (\bar{y}_i - g_{i,t}^{BAU} + z_i) \right)^2 \right. \\
 & \left. - C \sum_{j \in N} (g_{j,t}^{BAU} - z_j) \right\} + \delta^T v_i - \frac{b}{2} (\bar{y}_i - g_{i,1}^{BAU} + z_i - R_{i,1})^2 - \delta^T \frac{K}{2} \left(\frac{C}{K} \right)^2 \\
 & - C \sum_{j \in N} \left(g_{j,1}^{BAU} - z_j + R_{j,1} - \delta^T \frac{C}{K} \right). \tag{B1}
 \end{aligned}$$

The participants' pledges are given by maximizing $v_i + \omega \sum_{j \in M \setminus i} v_j$ from (B1) over z_i :

$$\begin{aligned}
 & \frac{\delta - \delta^T}{1 - \delta} \left[-b \left(\frac{K}{b+K} \right)^2 \left(\frac{C}{b} + \frac{C}{K} + z_i \right) - K \left(\frac{b}{b+K} \right)^2 \left(\frac{C}{b} + \frac{C}{K} + z_i \right) + \Omega(m)C \right] \\
 & - b \left(\frac{C}{b} + z_i \right) + \Omega(m)C = 0 \\
 \Leftrightarrow z_i = & [\Phi(m, T) - 1] \left(\frac{C}{b} + \frac{C}{K} \right) \in \left(\omega(m-1) \frac{C}{b}, \omega(m-1) \left(\frac{C}{b} + \frac{C}{K} \right) \right), \tag{B2}
 \end{aligned}$$

where $\Phi(m, T) \equiv 1 + \frac{\Omega(m)-1}{1 + \frac{1-\delta}{1-\delta^T} \frac{b}{K}} \in (1, \Omega(m))$. Substituting (B2) into (A2) and $d_{i,t} = \bar{y}_i - g_{i,t} - R_{i,t}$ for $g_{i,t} = g_{i,t}^{BAU} - z_i$, we get

$$R_{i,t} = \frac{b}{b+K} (\bar{y}_i - g_{i,t}^{BAU} + z_i) = \frac{b}{b+K} \Phi(\cdot) \left(\frac{C}{b} + \frac{C}{K} \right) = \Phi(\cdot) \frac{C}{K} \quad \forall t \in \{2, \dots, T\}, \tag{B3}$$

$$d_{i,t} = \bar{y}_i - g_{i,t}^{BAU} + z_i - R_{i,t} = \begin{cases} \Phi(\cdot) \frac{C}{b} + [\Phi(\cdot) - 1] \frac{C}{b} & \text{if } t = 1 \\ \Phi(\cdot) \left(\frac{C}{b} + \frac{C}{K} \right) - \Phi(\cdot) \frac{C}{K} = \Phi(\cdot) \frac{C}{b} & \text{if } t \in \{2, \dots, T\}. \end{cases} \tag{B4}$$

The next lemma specifies the optimal contract length for given coalition size m .

Lemma B1 *Let M^* denote an equilibrium coalition size $m^* \equiv |M^*|$. Then, a coalition of size $m = |M|$, satisfying $M \subseteq M^*$ or $M^* \subseteq M$, finds it optimal to contract for $T(m)$ periods,*

where

$$T(m) = \begin{cases} 1 & \text{if } m < \check{m}(x, \delta, m^*) \\ \{1, \dots, \infty\} & \text{if } m = \check{m}(x, \delta, m^*) \\ \infty & \text{if } m > \check{m}(x, \delta, m^*), \end{cases} \quad (\text{B5})$$

with

$$\check{m}(x, \delta, m^*) \equiv 1 + (m^* - 1) \sqrt{\frac{x+1}{x+2-\delta}} < m^*.$$

Proof of Lemma B1

We first prove that the optimal contract duration T^* of the equilibrium coalition M^* is infinity. If m^* countries contract for $T^* \leq \infty$ periods, each participant's continuation value from (6), (9) and (10) for $i \notin M$, and (B3) and (B4) for $i \in M$ is given by

$$\begin{aligned} v(m^*, T^*) &= \left\{ -\frac{b}{2} \left(\Phi(m^*, T^*) \frac{C}{b} + [\Phi(m^*, T^*) - 1] \frac{C}{K} \right)^2 - \delta \frac{K}{2} \left(\Phi(m^*, T^*) \frac{C}{K} \right)^2 \right. \\ &\quad \left. - C \left[\bar{y} - (m^* \Phi(m^*, T^*) + n - m^*) \left(\frac{C}{b} + \frac{\delta C}{K} \right) - m^* [\Phi(m^*, T^*) - 1] \frac{C}{K} \right] \right\} \\ &\quad + \sum_{t=2}^{T^*-1} \delta^{t-1} \left\{ -\frac{b}{2} \left(\Phi(m^*, T^*) \frac{C}{b} \right)^2 - \delta \frac{K}{2} \left(\Phi(m^*, T^*) \frac{C}{K} \right)^2 \right. \\ &\quad \left. - C \left[\bar{y} - (m^* \Phi(m^*, T^*) + n - m^*) \left(\frac{C}{b} + \frac{\delta C}{K} \right) \right] \right\} \\ &\quad + \delta^{T^*-1} \left\{ -\frac{b}{2} \left(\Phi(m^*, T^*) \frac{C}{b} \right)^2 - \delta \frac{K}{2} \left(\frac{C}{K} \right)^2 \right. \\ &\quad \left. - C \left[\bar{y} - (m^* \Phi(m^*, T^*) + n - m^*) \frac{C}{b} - n \frac{\delta C}{K} \right] \right\} + \delta^{T^*} v(m^*, T^*) \\ &= -\frac{1-\delta^{T^*}}{1-\delta} C \left\{ \bar{y} - C \left(m^* \Omega(m^*, T^*) - \frac{\Omega(m^*, T^*)^2}{2} + n - m^* \right) \left(\frac{1}{b} + \frac{1}{K} \right) \right. \\ &\quad \left. + \frac{C(1-\delta)(2n-1)}{2K} \right\} - \frac{C^2 [\Phi(m^*, T^*) - 1]^2 (b+K)}{2K^2} + \delta^{T^*} v(m^*, T^*) \end{aligned}$$

$$\begin{aligned}
&= -\frac{C}{1-\delta} \left\{ \bar{y} - C \left(m^* \Phi(m^*, T^*) - \frac{\Phi(m^*, T^*)^2}{2} + n - m^* \right) \left(\frac{1}{b} + \frac{1}{K} \right) \right. \\
&\quad \left. + \frac{C(1-\delta)(2n-1)}{2K} \right\} - \frac{C^2[\Phi(m^*, T^*) - 1]^2(b+K)}{2(1-\delta^{T^*})K^2}. \tag{B6}
\end{aligned}$$

Differentiating (B6) with respect to T^* yields

$$\begin{aligned}
\frac{\partial v(m^*, T^*)}{\partial T} &= \frac{\partial \Phi(m^*, T^*)}{\partial T} \frac{C^2[m^* - \Phi(m^*, T^*)](b+K)}{(1-\delta)bK} \\
&\quad - \frac{\partial \Phi(m^*, T^*)}{\partial T} \frac{C^2[\Phi(m^*, T^*) - 1](b+K)}{(1-\delta^{T^*})K^2} - \ln(\delta)\delta^{T^*} \frac{C^2[\Phi(m^*, T^*) - 1]^2(b+K)}{2(1-\delta^{T^*})^2K^2} \\
&= \frac{\partial \Phi(m^*, T^*)}{\partial T} \frac{C^2[m^* - \Omega(m^*)](b+K)}{(1-\delta)bK} - \ln(\delta)\delta^{T^*} \frac{C^2[\Phi(m^*, T^*) - 1]^2(b+K)}{2(1-\delta^{T^*})^2K^2}, \tag{B7}
\end{aligned}$$

where $\frac{\partial \Phi(m^*, T^*)}{\partial T} = -\frac{\ln(\delta)\delta^{T^*}}{1-\delta^{T^*}} \frac{\frac{1-\delta}{1-\delta^{T^*}} \frac{b}{K} [\Omega(m^*) - 1]}{\left(1 + \frac{1-\delta}{1-\delta^{T^*}} \frac{b}{K}\right)^2} \geq 0$, such that $T^* < \infty$ cannot be optimal.

$T^* = \infty$ raises the pledges and prevents the hold-up problem.

Now we derive the optimal contract duration T of a given coalition M . If m countries contract for T periods and $T^* = \infty$, each participant's continuation value from (6), (9) and (10) for $i \notin M$, and (B3) and (B4) for $i \in M$ is given by

$$\begin{aligned}
v(m, T) &= -\frac{1-\delta^T}{1-\delta} C \left\{ \bar{y} - C \left(m \Phi(m, T) - \frac{\Phi(m, T)^2}{2} + n - m \right) \left(\frac{1}{b} + \frac{1}{K} \right) \right. \\
&\quad \left. + \frac{C(1-\delta)(2n-1)}{2K} \right\} - \frac{C^2[\Phi(m, T) - 1]^2(b+K)}{2K^2} \\
&\quad - \frac{\delta^T}{1-\delta} C \left\{ \bar{y} - C \left(m^* \Phi(m^*, \infty) - \frac{\Phi(m^*, \infty)^2}{2} + n - m^* \right) \left(\frac{1}{b} + \frac{1}{K} \right) \right. \\
&\quad \left. + \frac{C(1-\delta)(2n-1)}{2K} \right\} - \delta^T \frac{C^2[\Phi(m^*, \infty) - 1]^2(b+K)}{2K^2}. \tag{B8}
\end{aligned}$$

Differentiating (B8) with respect to T yields

$$\frac{\partial v(m, T)}{\partial T} = \frac{\ln(\delta)\delta^T}{1-\delta} C \left\{ \bar{y} - C \left(m \Phi(m, T) - \frac{\Phi(m, T)^2}{2} + n - m \right) \left(\frac{1}{b} + \frac{1}{K} \right) \right.$$

$$\begin{aligned}
& \left. + \frac{C(1-\delta)(2n-1)}{2K} \right\} + (1-\delta^T) \frac{\partial \Phi(m,T)}{\partial T} \frac{C^2[m-\Omega(m)](b+K)}{(1-\delta)bK} \\
& - \frac{\ln(\delta)\delta^T}{1-\delta} C \left\{ \bar{y} - C \left(m^* \Phi(m^*, \infty) - \frac{\Phi(m^*, \infty)^2}{2} + n - m^* \right) \left(\frac{1}{b} + \frac{1}{K} \right) \right. \\
& \left. + \frac{C(1-\delta)(2n-1)}{2K} \right\} - \ln(\delta)\delta^T \frac{C^2[\Phi(m,T)-1]^2(b+K)}{2K^2} \\
& = (1-\delta^T) \frac{\partial \Phi(m,T)}{\partial T} \frac{C^2[m-\Omega(m)](b+K)}{(1-\delta)bK} - \ln(\delta)\delta^T \frac{C^2[\Phi(m^*, \infty)-1]^2(b+K)}{2K^2} \\
& - \frac{\ln(\delta)\delta^T C^2}{1-\delta} \left(m\Phi(m,T) - \frac{\Phi(m,T)^2}{2} - m \right) \left(\frac{1}{b} + \frac{1}{K} \right) \\
& + \frac{\ln(\delta)\delta^T C^2}{1-\delta} \left(m^* \Phi(m^*, \infty) - \frac{\Phi(m^*, \infty)^2}{2} - m^* \right) \left(\frac{1}{b} + \frac{1}{K} \right). \tag{B9}
\end{aligned}$$

Substituting the definitions of $\Phi(m, T)$ and $\Omega(m)$ into (B9) and simplifying yields

$$\frac{\partial v(m,T)}{\partial T} = - \frac{\ln(\delta)\delta^T \omega(2-\omega)C^2(x+1) \left(x + 2\frac{1-\delta}{1-\delta^T} \right)}{2b(1-\delta) \left(x + \frac{1-\delta}{1-\delta^T} \right)^2} [(m-1)^2 - (m^*-1)^2 \Lambda(T)], \tag{B10}$$

with

$$\Lambda(T) \equiv \frac{\left(x + \frac{1-\delta}{1-\delta^T} \right)^2}{(x+1-\delta) \left(x + 2\frac{1-\delta}{1-\delta^T} \right)} \quad \text{with} \quad \frac{\partial \Lambda(T)}{\partial T} = \frac{2 \ln(\delta)\delta^T (1-\delta)^2 \left(x + \frac{1-\delta}{1-\delta^T} \right)}{(1-\delta^T)^3 (x+1-\delta) \left(x + 2\frac{1-\delta}{1-\delta^T} \right)^2} < 0.$$

$\frac{\partial \Lambda(T)}{\partial T} < 0$ implies that $\frac{\partial v(m,T)}{\partial T}$ is positive [negative] for all T if it is positive [negative] for $T = 1$ [$T = \infty$]. By contrast, if $\frac{\partial v(m,T)}{\partial T}$ is negative for $T = 1$ but positive for $T = \infty$, it is minimal for some $T \in (1, \infty)$ and it is either maximal for $T = 1$ or for $T = \infty$. Thus, the optimal contract duration T of a given coalition M is either $T = 1$ or $T = \infty$. From (B8), we get the difference in each participant's continuation value between $T = 1$ and $T = \infty$

$$\begin{aligned}
v(m, 1) - v(m, \infty) &= -C \left\{ \bar{y} - C \left(m\Phi(m, 1) - \frac{\Phi(m, 1)^2}{2} + n - m \right) \left(\frac{1}{b} + \frac{1}{K} \right) \right. \\
& \left. + \frac{C(1-\delta)(2n-1)}{2K} \right\} - \frac{C^2[\Phi(m, 1)-1]^2(b+K)}{2K^2}
\end{aligned}$$

$$\begin{aligned}
& -\frac{\delta}{1-\delta}C \left\{ \bar{y} - C \left(m^* \Phi(m^*, \infty) - \frac{\Phi(m^*, \infty)^2}{2} + n - m^* \right) \left(\frac{1}{b} + \frac{1}{K} \right) \right. \\
& \left. + \frac{C(1-\delta)(2n-1)}{2K} \right\} - \delta \frac{C^2[\Phi(m^*, \infty) - 1]^2(b+K)}{2K^2} \\
& + \frac{C}{1-\delta} \left\{ \bar{y} - C \left(m \Phi(m, \infty) - \frac{\Phi(m, \infty)^2}{2} + n - m \right) \left(\frac{1}{b} + \frac{1}{K} \right) \right. \\
& \left. + \frac{C(1-\delta)(2n-1)}{2K} \right\} + \frac{C^2[\Phi(m, \infty) - 1]^2(b+K)}{2K^2}. \tag{B11}
\end{aligned}$$

Substituting the definitions of $\Phi(m, T)$ and $\Omega(m)$ into (B11) and simplifying yields

$$v(m, 1) - v(m, \infty) = -\frac{\delta\omega(2-\omega)C^2}{2b(1-\delta)} \frac{x+2-\delta}{x+1-\delta} \left[(m-1)^2 - (m^*-1)^2 \frac{x+1}{x+2-\delta} \right]. \tag{B12}$$

From (B12), the optimal contract duration is one period if $m < \check{m}(x, \delta, m^*)$, infinity if $m > \check{m}(x, \delta, m^*)$, and arbitrary if $m = \check{m}(x, \delta, m^*)$, where $\check{m}(x, \delta, m^*)$ is defined in Lemma B1. QED

Defining

$$\check{m}_M(x, \delta) \equiv 1 + \frac{1}{1 - \sqrt{\frac{x+1}{x+2-\delta}}}, \quad \check{m}_{\bar{I}}(x, \delta, \omega) \equiv 1 + \frac{2/\omega}{1 - \frac{2-\omega}{\omega} \frac{\delta}{x+1-\delta}} > m_{\bar{I}}(\omega), \tag{B13}$$

the next proposition specifies the equilibrium coalition size m^* . Note that $\check{m}_{\bar{I}} < \hat{m}_{\bar{I}}$ and

$$\check{m}_M > m_M \Leftrightarrow \frac{x+1}{x+2-\delta} > \frac{x+\delta}{x+1} \Leftrightarrow \frac{(1-\delta)^2}{(x+2-\delta)(x+1)} > 0,$$

such that it is more likely that the incentive constraint is binding with time-independent pledges than with time-dependent pledges.

Proposition B1 *M^* is an equilibrium coalition if and only if either $m^* = \lfloor m_{\bar{I}}(\omega) \rfloor$ or $\lfloor m_{\bar{I}}(\omega) \rfloor < m^* \leq \min\{n, \check{m}(x, \delta, \omega)\}$, where*

$$\check{m}(x, \delta, \omega) = \min\{\check{m}_M(x, \delta), \check{m}_{\bar{I}}(x, \delta, \omega)\} = \begin{cases} \check{m}_M(x, \delta) & \text{if } x < \check{x}(\delta, \omega) \\ \check{m}_{\bar{I}}(x, \delta, \omega) & \text{if } x \geq \check{x}(\delta, \omega), \end{cases} \tag{B14}$$

with

$$\check{x}(\delta, \omega) \equiv \hat{x}(\delta, \omega) - (1 - \delta) \leq \frac{2 - \omega}{\omega}, \quad \frac{\partial \check{x}}{\partial \delta} > 0, \quad \frac{\partial \check{x}}{\partial \omega} < 0,$$

where $\delta \geq 0.36 \vee \omega \leq 0.59 \implies \check{x}(\delta, \omega) > 0$.

Proof of Proposition B1

We first derive the condition for external stability. If a non-participant joins in equilibrium, then $m = m^* + 1$, which is not beneficial to him if his continuation value in case of participation for $m = m^* + 1$ from (B6) for $T^* = \infty$ falls short of his continuation value in case of non-participation for $m = m^*$ from (6) for $d_{i,t}$ and $R_{i,t+1}$ from (9) and (10):

$$\begin{aligned} & -\frac{C}{1-\delta} \left\{ \bar{y} - C \left((m^* + 1)\Phi(m^* + 1, \infty) - \frac{\Phi(m^* + 1, \infty)^2}{2} + n - (m^* + 1) \right) \left(\frac{1}{b} + \frac{1}{K} \right) \right. \\ & \left. + \frac{C(1-\delta)(2n-1)}{2K} \right\} - \frac{C^2[\Phi(m^* + 1, \infty) - 1]^2(b+K)}{2K^2} < \\ & -\frac{C}{1-\delta} \left[\bar{y} - C \left(m^*\Phi(m^*, \infty) - \frac{1}{2} + n - m^* \right) \left(\frac{1}{b} + \frac{\delta}{K} \right) \right] + \frac{C^2 m^* [\Phi(m^*, \infty) - 1]}{K} \\ \Leftrightarrow & \left[(m^* + 1)\Phi(m^* + 1, \infty) - \frac{\Phi(m^* + 1, \infty)^2}{2} - m^*\Phi(m^*, \infty) - \frac{1}{2} \right] \frac{C^2}{1-\delta} \left(\frac{1}{b} + \frac{\delta}{K} \right) \\ & - \frac{(1-\delta)[\Phi(m^* + 1, \infty) - 1]^2}{2x} \frac{C^2}{1-\delta} \left(\frac{1}{b} + \frac{\delta}{K} \right) < 0 \\ \Leftrightarrow & -m^*\omega^2 \left(m^* - \frac{2}{\omega} \right) \frac{C^2}{2(1-\delta)} \left(\frac{1}{b} + \frac{\delta}{K} \right) \frac{x}{x+1-\delta} < 0, \end{aligned} \tag{B15}$$

requiring $m^* > 2/\omega$ for external stability.

Now we derive the conditions for internal stability. If a single country deviates by not participating, the remaining coalition sets $T = 1$ only if $m^* - 1 \leq \check{m}(x, \delta, m^*) \Leftrightarrow m^* \leq \check{m}_M(x, \delta)$ as defined in Proposition B1. Suppose $m^* > \check{m}_M(x, \delta)$. If a participant deviates in equilibrium, then $m = m^* - 1 > \check{m}(x, \delta, m^*)$; so $T = \infty$ by Lemma B1. Such a permanent deviation is not beneficial to him if his continuation value in case of participation for $m = m^*$ from (B6) for $T^* = \infty$ exceeds his continuation value in case of non-participation

for $m = m^* - 1$ from (6) for $d_{i,t}$ and $R_{i,t+1}$ from (9) and (10):

$$\begin{aligned}
& -\frac{C}{1-\delta} \left\{ \bar{y} - C \left(m^* \Phi(m^*, \infty) - \frac{\Phi(m^*, \infty)^2}{2} + n - m^* \right) \left(\frac{1}{b} + \frac{1}{K} \right) \right. \\
& \left. + \frac{C(1-\delta)(2n-1)}{2K} \right\} - \frac{C^2[\Phi(m^*, \infty) - 1]^2(b+K)}{2K^2} \geq \\
& -\frac{C}{1-\delta} \left[\bar{y} - C \left((m^* - 1)\Phi(m^* - 1, \infty) - \frac{1}{2} + n - (m^* - 1) \right) \left(\frac{1}{b} + \frac{\delta}{K} \right) \right] \\
& + \frac{C^2(m^* - 1)[\Phi(m^* - 1, \infty) - 1]}{K} \\
\Leftrightarrow & \left[m^* \Phi(m^*) - \frac{\Phi(m^*)^2}{2} - (m^* - 1)\Phi(m^* - 1) - \frac{1}{2} \right] \frac{C^2}{1-\delta} \left(\frac{1}{b} + \frac{\delta}{K} \right) \\
& - \frac{(1-\delta)[\Phi(m^*, \infty) - 1]^2}{2x} \frac{C^2}{1-\delta} \left(\frac{1}{b} + \frac{\delta}{K} \right) \geq 0 \\
\Leftrightarrow & (m^* - 1)\omega^2 \left(\frac{2}{\omega} - (m^* - 1) \right) \frac{C^2}{2(1-\delta)} \left(\frac{1}{b} + \frac{\delta}{K} \right) \frac{x}{x+1-\delta} \geq 0, \tag{B16}
\end{aligned}$$

requiring $m^* \leq m_I(\omega) = 1 + 2/\omega$ for internal stability. From (B15), the coalition is externally stable if $m^* > 2/\omega$, which is only fulfilled for the largest internally stable coalition. Thus, $m^* > \check{m}_M(x, \delta)$ implies $m^* = \lceil 1 + 2/\omega \rceil$.

Now suppose $m^* \leq \check{m}_M(x, \delta)$. If a participant deviates in equilibrium, then $m = m^* - 1 \leq \check{m}(x, \delta, m^*)$; so $T = 1$ by Lemma 2, and the participant is expected to join the coalition next period. Such a one-period deviation is not beneficial to him if his one-period utility in case of participation for $m = m^*$ from (B6) for $T^* = \infty$ exceeds his one-period utility in case of non-participation for $m = m^* - 1$ from (6) for $d_{i,t}$ and $R_{i,t+1}$ from (9) and (10):

$$\begin{aligned}
& -C \left\{ \bar{y} - C \left(m^* \Phi(m^*, \infty) - \frac{\Phi(m^*, \infty)^2}{2} + n - m^* \right) \left(\frac{1}{b} + \frac{\delta}{K} \right) \right. \\
& \left. + \frac{C(1-\delta)(2n-1)}{2K} \right\} - (1-\delta) \frac{C^2[\Phi(m^*, \infty) - 1]^2(b+K)}{2K^2} \geq \\
& -\frac{b}{2} \left(\frac{C}{b} \right)^2 - \delta \frac{K}{2} \left(\frac{C}{K} \right)^2 + \frac{C^2(m^* - 1)[\Phi(m^* - 1, 1) - 1]}{K}
\end{aligned}$$

$$\begin{aligned}
& -C \left[\bar{y} - ((m^* - 1)\Phi(m^* - 1, 1) + n - (m^* - 1)) \frac{C}{b} - n \frac{\delta C}{K} \right] \\
\Leftrightarrow & \left[m^* \Phi(m^*, \infty) - \frac{\Phi(m^*, \infty)^2}{2} - (m^* - 1)\Phi(m^* - 1, 1) - \frac{1}{2} \right] \frac{1+x}{x} \frac{C^2}{b} \\
& - \frac{(1-\delta)[\Phi(m^*, \infty) - 1]^2}{2} \frac{\delta}{x^2} \frac{C^2}{b} \geq 0 \\
\Leftrightarrow & (m^* - 1)\omega^2 \left[\frac{2}{\omega} + (m^* - 1) \left(\frac{2-\omega}{\omega} \frac{\delta}{x+1-\delta} - 1 \right) \right] \frac{C^2}{2b} \geq 0, \tag{B17}
\end{aligned}$$

requiring $m^* \leq \check{m}_{\bar{I}}(x, \delta, \omega) = 1 + \frac{2/\omega}{1 - \frac{2-\omega}{\omega} \frac{\delta}{x+1-\delta}}$ for internal stability. For $x \leq \frac{2\delta-\omega}{\omega}$, any coalition is internally stable. From (A14), the coalition is externally stable if $m^* > 2/\omega$, which is definitely fulfilled for the largest internally stable coalition. Furthermore, $m^* = \check{m}_M(x, \delta) < \lfloor 1 + 2/\omega \rfloor$ would imply $m^* = \check{m}_M(x, \delta) \leq 2/\omega$, such that the coalition would not be externally stable and $m^* > \check{m}_M(x, \delta)$ would hold. Thus, $m^* \leq \check{m}_M(x, \delta)$ implies $m^* \in \left[\lfloor 1 + 2/\omega \rfloor, \min\{n, \check{m}_{\bar{I}}(x, \delta, \omega)\} \right]$.

Comparing $\check{m}_M(x, \delta)$ with $\check{m}_{\bar{I}}(x, \delta, \omega)$ yields $\check{m}_M(x, \delta) \gtrless \check{m}_{\bar{I}}(x, \delta, \omega) \Leftrightarrow x \gtrless \check{x}(\delta, \omega)$ as defined in Proposition B1, which proves the cases in (B14). Differentiating $\check{x}(\delta, \omega)$ from Proposition 1 with respect to δ and ω , we get $\frac{\partial \check{x}}{\partial \delta} = \frac{\partial \hat{x}}{\partial \delta} + 1$ and $\frac{\partial \check{x}}{\partial \omega} = \frac{\partial \hat{x}}{\partial \omega}$, which proves the signs of the derivatives in Proposition B1. Finally,

$$\check{x}(\delta, 1) = \frac{[7 + \sqrt{17} - \delta][\delta - 7 + \sqrt{17}]}{5 - 7\delta + \sqrt{1 + 14\delta + \delta^2}} \gtrless 0 \quad \Leftrightarrow \quad \delta \gtrless 0.36, \tag{B18}$$

$$\check{x}(0, \omega) = \frac{[2 + \sqrt{2} - \omega][2 - \sqrt{2} - \omega]}{0.5\omega(4 - \omega)} \gtrless 0 \quad \Leftrightarrow \quad \omega \gtrless 0.59 \tag{B19}$$

and $\check{x}(1, \omega) = (2 - \omega)/\omega$ proves the bounds of $\check{x}(\delta, \omega)$ in Proposition B1. QED

The next proposition compares the welfare of a coalition country with the Kyoto Protocol and with the Paris Agreement.

Proposition B2 *In economies $\check{\mathcal{E}}_2$ and $\check{\mathcal{E}}_3$ [$\check{\mathcal{E}}_1$] the welfare of a coalition country is higher with the Kyoto Protocol ($m^* = 37, \omega = 1$) than with the Paris Agreement ($m^* = 195, \omega = \frac{1}{97}$) [if and only if $\omega < 0.0174$].*

To prove Proposition B2, we first characterize the feasible economies $\check{\mathcal{E}}_1$ - $\check{\mathcal{E}}_3$ in Lemma B2 and then derive the welfare difference of each participant and that of each non-participant between the Kyoto Protocol and the Paris Agreement in Lemma B3.

Lemma B2 *Suppose $m^* = 37$ holds for $\omega = 1$ (Kyoto Protocol) and $m^* = 195$ holds for $\omega \leq 0.5$ (Paris Agreement). Then, Table 5 characterizes the feasible economies.*

Economy	Paris Agreement	Kyoto Protocol	$x \in$	$\delta \in$	$\omega \in$
$\check{\mathcal{E}}_1$	$m^* = \check{m}_M(x, \delta)$	$m^* = \check{m}_I(x, \delta, 1)$	1.016	0.979	[0.01, 0.5)
$\check{\mathcal{E}}_2$	$m^* = m_I(\omega)$	$m^* = \check{m}_I(x, \delta, 1)$	(0.839, 1.059]	(0.893, 1]	0.01
$\check{\mathcal{E}}_3$	$m^* = m_I(\omega)$	$m^* = \check{m}_M(x, \delta)$	(0, 0.839]	[0.893, 0.942)	0.01

Table B1: Feasible economies $\check{\mathcal{E}}_1$ - $\check{\mathcal{E}}_3$

Proof of Lemma B2

For the Kyoto Protocol, $m^* = 37$ and $\omega = 1$ imply that either $m^* \leq \check{m}_M(x, \delta)$ or $m^* \leq \check{m}_I(x, \delta, \omega = 1)$ is binding. Else, $m^* = m_I(\omega = 1) = 3$ or $m^* = n = 197$ would hold. First suppose $m^* \leq \check{m}_M(x, \delta)$ and $m^* = \check{m}_I(x, \delta, \omega = 1)$ hold for the Kyoto Protocol. From (B13), we then get

$$\check{m}_M(x, \delta) = 1 + \frac{1}{1 - \sqrt{\frac{x+1}{x+2-\delta}}} \geq 37 \quad \Leftrightarrow \quad \delta \geq \frac{1154 - 71x}{1225}, \quad (\text{B20})$$

$$\check{m}_I(x, \delta, \omega = 1) = 1 + \frac{2}{1 - \frac{\delta}{x+1-\delta}} = 37 \quad \Leftrightarrow \quad x = \frac{35\delta - 17}{17}. \quad (\text{B21})$$

Substituting (B21) into (B20) and rearranging yields $\delta \geq \frac{595}{666} \approx 0.893$, and substituting this into (B21) yields $x \geq \frac{559}{666} \approx 0.839$. $\delta \leq 1$ then implies $\delta \in [\frac{595}{666}, 1]$ and $x \in [\frac{559}{666}, \frac{18}{17}]$.

For the Paris Agreement, $m^* = 195$ and $\omega \leq 0.5$ imply that $m^* \leq m_I(\omega)$, $m^* \leq \check{m}_M(x, \delta)$ or $m^* \leq \check{m}_I(x, \delta, \omega)$ is binding. Else, $m^* = n = 197$ would hold. Substituting $x = \frac{35\delta - 17}{17}$ from (B21) into (B13), we get

$$\check{m}_M\left(x = \frac{35\delta - 17}{17}, \delta\right) = 1 + \frac{1}{1 - \sqrt{\frac{35\delta}{18\delta + 17}}} = 195 \quad \Leftrightarrow \quad \delta = \frac{633233}{646778} \approx 0.979, \quad (\text{B22})$$

$$\check{m}_{\bar{I}}\left(x = \frac{35\delta - 17}{17}, \delta, \omega\right) = 1 + \frac{36}{35\omega - 34} = 195 \quad \Leftrightarrow \quad \omega = \frac{3316}{3395} \approx 0.977. \quad (\text{B23})$$

$\check{m}_{\bar{I}}\left(x = \frac{35\delta - 17}{17}, \delta, \omega\right) = 195$ cannot hold since $\omega < 0.5$, and $\check{m}_M\left(x = \frac{35\delta - 17}{17}, \delta\right) = 195$ holds for $\delta \approx 0.979$ and $x = \frac{35\delta - 17}{17} \approx 1.016$. Finally, $m_{\bar{I}}(\omega) = 1 + \frac{2}{\omega} = 195$ holds for $\omega = \frac{1}{97} \approx 0.010$. Thus, economies $\check{\mathcal{E}}_1$ are characterized by $\check{m}_{\bar{I}}(x, \delta, \omega = 1) = 37$, $\check{m}_M(x, \delta) = 195$ and $m_{\bar{I}}(\omega) \geq 195$, which implies the values in the second line of Table 5, and economies $\check{\mathcal{E}}_2$ are characterized by $\check{m}_{\bar{I}}(x, \delta, \omega = 1) = 37$, $\check{m}_M(x, \delta) \geq 37$ and $\check{m}_{\bar{I}}(\omega) = 195$, which implies the values in the third line of Table 5.

Now suppose $m^* = \check{m}_M(x, \delta)$ and $m^* \leq \check{m}_{\bar{I}}(x, \delta, \omega = 1)$ hold for the Kyoto Protocol. From (B13), we then get

$$\check{m}_M(x, \delta) = 1 + \frac{1}{1 - \sqrt{\frac{x+1}{x+2-\delta}}} = 37 \quad \Leftrightarrow \quad \delta = \frac{1154 - 71x}{1225}, \quad (\text{B24})$$

$$\check{m}_{\bar{I}}(x, \delta, \omega = 1) = 1 + \frac{2}{1 - \frac{\delta}{x+1-\delta}} \geq 37 \quad \Leftrightarrow \quad x \leq \frac{35\delta - 17}{17}. \quad (\text{B25})$$

Substituting (B25) into (B24) and rearranging yields $\delta \geq \frac{595}{666} \approx 0.893$, and substituting this into (B25) yields $x \leq \frac{559}{666} \approx 0.839$. $x \geq 0$ then implies $\delta \in \left[\frac{595}{666}, \frac{1154}{1225}\right]$ and $x \in \left[0, \frac{559}{666}\right]$.

For the Paris Agreement, $m^* = 195$ and $\omega \leq 0.5$ imply that $m^* \leq m_{\bar{I}}(\omega)$ or $m^* \leq \check{m}_{\bar{I}}(x, \delta, \omega)$ is binding. Else, $\check{m}_M(x, \delta) = 37$ or $m^* = n = 197$ would hold. Substituting $x = \frac{1154 - 1225\delta}{71}$ from (B24) into (B13), we get

$$\check{m}_{\bar{I}}\left(x = \frac{1154 - 1225\delta}{71}, \delta, \omega\right) = 1 + \frac{142\delta - 2450(1 - \delta)}{142\delta - 1225\omega(1 - \delta)} = 195 \quad \Leftrightarrow \quad \omega = \frac{12478\delta + 1225}{118825(1 - \delta)}. \quad (\text{B26})$$

$\check{m}_{\bar{I}}\left(x = \frac{1154 - 1225\delta}{71}, \delta, \omega\right) = 195$ cannot hold since $\delta \geq \frac{595}{666}$ implies $\omega \geq \frac{3316}{3395} > 0.5$. Finally, $m_{\bar{I}}(\omega) = 1 + \frac{2}{\omega} = 195$ holds for $\omega = \frac{1}{97} \approx 0.010$. Thus, economies $\check{\mathcal{E}}_3$ are characterized by $\check{m}_M(x, \delta) = 37$, $\check{m}_{\bar{I}}(x, \delta, \omega = 1) \geq 37$ and $m_{\bar{I}}(\omega) = 195$, which implies the values in the last line of Table 5. QED

Lemma B3 *Suppose $m^* = 37$ holds for $\omega = 1$ (Kyoto Protocol) and $m^* = 195$ holds for*

$\omega \leq 0.5$ (*Paris Agreement*). Then, the welfare of each participant [non-participant] is higher with the Kyoto Protocol than with the Paris Agreement if and only if $\omega < 0.0174$ [$\omega < 0.0352$]. Furthermore, the intertemporal climate damage is smaller with the Kyoto Protocol than with the Paris Agreement if and only if $\omega < 0.0352$.

Proof of Lemma B3

For the Kyoto Protocol, we have $m^* = 37$ and $\omega = 1$, and for the Paris Agreement, we have $m^* = 195$ and $\omega \leq 0.5$. From (B6) for $T^* = \infty$, we get the welfare difference of each participant

$$v(m = 37, \omega = 1) - v(m = 195, \omega < 1) = \frac{C^2(x + 1)[648 - 18818\omega(2 - \omega)]}{b(1 - \delta)(x + 1 - \delta)}, \quad (\text{B27})$$

which is positive [negative] for $\omega < [>]0.0174$, and from the first line's right-hand side of (B15), we get the welfare difference of each non-participant

$$v(m = 37, \omega = 1) - v(m = 195, \omega < 1) = \frac{C^2(x + 2 - \delta)[1332 - 37830\omega]}{b(1 - \delta)(x + 1 - \delta)}, \quad (\text{B28})$$

which is positive [negative] for $\omega < [>]0.0352$. Since each nonparticipant always chooses the business-as-usual energy consumption and technology investment, its welfare difference stems from the difference in the intertemporal climate damage, such that $\omega < [>]0.0352$ implies a smaller [greater] intertemporal climate damage with the Kyoto Protocol than with the Paris Agreement. QED

Proof of Proposition B2

From Lemma B2, $\omega = 0.1$ [$\omega \in [0.1, 0.5]$] holds in economies $\check{\mathcal{E}}_2$ and $\check{\mathcal{E}}_3$ [$\check{\mathcal{E}}_1$], such that the welfare of each participant is higher with the Kyoto Protocol than with the Paris Agreement in economies $\check{\mathcal{E}}_2$ and $\check{\mathcal{E}}_3$ [if and only if $\omega < 0.0174$ in economies $\check{\mathcal{E}}_1$] from Lemma B3. QED