Voluntary Participation in International Environmental Agreements and Authority Structures in a Federation: A Note *

Ryusuke Shinohara†

February 19, 2021
Forthcoming in *Environmental and Resource Economics*

Abstract

We examine how a voluntary participation decision in international environmental negotiations affects the endogenous authority structure in a federation. In our model, the federal government of each country decides whether to delegate both the decision to participate in a negotiation that determines the abatement level of pollution (the level of the public good), and the negotiation itself, to a regional government of the polluter region. We show that there exists a subgame perfect equilibrium in which none of the federal governments chooses delegation, which is quite different from the authority structure in the absence of a voluntary participation decision. The main contribution is to explain why the federal government has an incentive not to delegate decisions to a regional government.

Keywords: Delegation; International environmental agreements; Nash bargaining; Negotiation; Participation; Public goods.
JEL classification: C78, D62, H41, H77.

*I thank the Co-editor, Michael Finus, and the anonymous reviewers for their helpful comments. I am grateful to Kimiko Terai and Masayoshi Hayashi for their helpful suggestions. In addition, I thank the participants of the 76th meeting of the Japan Institution of Public Finance, and the DC conference in Japan. I gratefully acknowledge the financial support from KAKENHI Grant-in-Aid for Scientific Research (B) (No. 19H01483) and (C) (No. 18K01519). The usual disclaimer applies.

†Correspondence: Ryusuke Shinohara, Department of Economics, Hosei University, 4342, Aihara-machi, Machida-shi, Tokyo, 194-0298, Japan. E-mail address: ryusukes@hosei.ac.jp

1 Introduction

Various attempts have been made to solve transboundary environmental problems through international environmental agreements (IEAs) between the countries involved. As readily observed from real-world evidence, including on IEAs concerning the abatement of greenhouse gases, the effectiveness of many IEAs is subject to the voluntary participation of the countries involved. In federations such as the United States and Canada, or in political unions like the European Union, the authority for environmental regulations is sometimes granted to lower-tier governments; for example, those at the regional level.[1] As a result, the prevailing authority structure in a federation will have some effect on the negotiations of IEAs. Given this background, we examine how the decision to voluntarily participate in IEAs and the authority structure in a federation influence each other.

This question is new to the literature. Studies such as Carraro and Siniscalco (1993), Barrett (1994), and Rubio and Ulph (2006) investigate the voluntary participation property of IEAs but are largely unconcerned about the authority structure in a federation. Similarly, Eckert (2003) and Buchholz et al. (2013) examine how the distribution of authority affects IEAs, but only *in the absence of* a voluntary participation decision by the countries involved.

To respond to our question, we construct a simple model with n identical federal countries, based on the models in Eckert (2003) and Buchholz et al. (2013). Each country consists of two autonomous regions: one a polluter region and the other a nonpolluter region. Every country is governed by a federal government (FG) and each of the subnational regions is governed by a regional government (RG). There is then an opportunity for international negotiation, which aims to internalize the externalities relating to pollution abatement. In addition, and in contrast to Eckert (2003) and Buchholz et al. (2013), we introduce a stage in which each country decides whether it participates in the negotiation. Thus, for the FG, the delegation choice includes a participation decision. That is, the FG of each country decides whether to delegate both the participation decision and the negotiation to the polluter RG.

Interestingly, the equilibrium delegation structure in a federation is completely different with and without voluntary participation. Without the participation decision in a two-country model, Eckert (2003) shows that an equilibrium is supported in which all FGs delegate the negotiation to the polluter region if the region is sufficiently populous. In contrast, there is no equilibrium in which all FGs themselves negotiate if countries are identical. These results carry over to the model with many identical countries. However, with the participation decision, Proposition 1 shows that an equilibrium always exists such that all FGs delegate neither the participation decision nor the negotiation.

Our result contributes to explaining why the FG itself makes the participation decision and negotiates, which is required because, at least in the real world, the national government decides some IEAs, such as those concerning climate change. However, existing studies on the authority structure in a federation (Eckert, 2003; Buchholz et al., 2013) and on strategic delegation (e.g., Segendorff, 1998) have not pointed out that there is a sufficient possibility that nondelegation takes place. Our analysis reveals that the strategic advantage of delegation in a negotiation, which has been reported in these existing studies, decreases in the presence of the participation decision and, hence, the FG chooses nondelegation (see the discussion following Proposition 1).

In addition, in studies on fiscal federalism (e.g., Besley and Coate, 2003), the centralized provision of a public good is preferable if the spillovers associated with the good are sufficiently strong. Our result thus provides new insight into why the FG chooses centralization ("nondelegation" in our terminology) in relation to the voluntary participation behavior, but not the strength of the spillovers .

The remainder of the paper is organized as follows. Section 2 introduces the model. Section 3 provides the results. Section 4 discusses some extensions.

[1] See Dalmazzone (2006) for a survey article on environmental federalism.

2 The model

Assume $N = \{1, \dots, n\}$ is a set of *identical* countries such that $n \geq 3$. Each country consists of two regions. One of the regions is a "polluter" region and the other is a "nonpolluter" region. The population of each country is normalized to unity. In every country, α $(0 < \alpha < 1)$ portion of residents resides in the polluter region and $1 - \alpha$ resides in the nonpolluter region.

The polluter region invests in pollution reduction, which is a *public good*. The right to decide on the pollution reduction level is assigned to either the FG of the country or the RG of the polluter region, as described in detail later. The nonpolluter region is never assigned this right of control over the decision. The level of pollution reduction by country i is denoted by $g_i \geq 0$ and the cost of reducing the pollution is measured by a quadratic function $c(g_i) = \frac{(g_i)^2}{2}$.

Each country $i \in N$ has a FG, which has the objective of maximizing the country's total surplus $v(\sum_{j \in N} g_j) - c(g_i) + T_{i,F}$. Country i reaps a benefit from its own pollution reduction and from the pollution reduction of other countries. For the pollution reduction levels of $(g_j)_{j \in N}$, country i's benefit is measured by a linear benefit function $v(\sum_{j \in N} g_j) = \sum_{j \in N} g_j$. $T_{i,F}$ represents international transfers from other countries to the FG of country i.

The polluter region of country $i \in N$ is governed by a RG, which has the objective of maximizing the regional surplus $\alpha v(\sum_{j \in N} g_j) - c(g_i) + T_{i,R}$. We assume that the polluter region has the technology to reduce the pollution and that it incurs the associated costs. The RG is only concerned with the benefit of its residents and hence, the benefit term becomes $\alpha v(\sum_{j \in N} g_j) = \alpha \sum_{j \in N} g_j$. $T_{i,R}$ represents the international transfers that the RG receives from other countries.

The FG of each country then determines whether to delegate the decision concerning participation in the environmental negotiation and the negotiation itself to the RG of its polluter region. Formally, we consider the following multistage game with complete information.

Stage 1 The FG of each country simultaneously decides whether to delegate all decisions in Stages 2 and 3 to the polluter region's government.[2] The set of possible actions of each FG is denoted by $\{F, R\}$, where F denotes that the FG itself makes all decisions in Stages 2 and 3 and R is that the polluter region's RG does so. All decisions in subsequent stages are made by the RG *in the best interests of the region* if the FG chooses R.

Stage 2 Each country decides whether it will participate in the negotiation held in Stage 3. If the FG of a country chooses R in Stage 1, then the participation decision is made by the polluter RG. Otherwise, the decision is made by the FG. The set of possible actions of each country is denoted by $\{I, O\}$, where I denotes "IN" (participation) and O denotes "OUT" (nonparticipation).

Stage 3 The participants negotiate the pollution reduction levels and the international transfers for each participant. This negotiation is analyzed by the *Nash bargaining solution* (NBS). The negotiation outcome reflects the negotiators' preferences. That is, if the FG of a country chooses R in Stage 1, then the negotiation outcome reflects the preferences of the RG. Otherwise, the outcome reflects the FG's preferences. A nonparticipating country reduces its pollution independently from the negotiation, and it neither receives nor makes any transfers. If the negotiation breaks down, then every delegate independently reduces pollution without transfers.

Let $n_R \in \{0, \dots, n\}$ be the number of FGs that choose R in Stage 1. We characterize actions taken in Stage 1 by n_R because the number of FGs choosing F is automatically decided by $n - n_R$. Similarly, let $(p, p_R) \in \{0, \dots, n\} \times \{0, \dots, p\}$, where p is the number of countries that choose I in Stage 2 and p_R is the number of participating RGs. We characterize a set of participants $P \subseteq N$ by the pair (p, p_R)

[2] In Section 4, we discuss the effect of intermediate delegations as an extension.

because the number of participating countries with FGs that choose F in Stage 1 is automatically determined by $p - p_R$. We examine the *subgame perfect Nash equilibrium* (SPNE) of this model.

The assumptions that every country has the linear benefit function and the quadratic cost function are frequently used in studies of voluntary participation games (e.g., Barrett, 2006; Finus and Pintassilgo, 2012; Goeschl and Perino, 2017).[3]

3 Results

We provide a summary of the Nash bargaining outcome for a set of participants $P \subseteq N$ with (p, p_R).[4] We introduce a parameter α_i such that $\alpha_i = 1$ if country i's second-stage player is the FG and $\alpha_i = \alpha$ if it is the RG of the polluter region.

If negotiation breaks down, then the negotiator of country $i \in P$ determines the pollution reduction level to maximize its payoff. Let g_i^d be the pollution reduction level of country i if negotiation breaks down. Then, $g_i^d = \alpha \equiv g_R^d$ if $\alpha_i = \alpha$ and $g_i^d = 1 \equiv g_F^d$ if $\alpha_i = 1$. In the negotiation, the participants in P negotiate the pollution reduction levels, denoted by $(g_j^P)_{j\in P}$, and the transfers, denoted by $(T_j^P)_{j\in P}$. Through the Nash bargaining, $(g_j^P)_{j\in P}$ maximizes the participants' surplus relative to the breakdown of the negotiation $\mathcal{S}(P) \equiv \sum_{i\in P}[\alpha_i v(\sum_{j\in P} g_j^P + \sum_{j\in N\setminus P} g_j^d) - c(g_i^P)] - \sum_{i\in P}[\alpha_i v(\sum_{j\in N} g_j^d) - c(g_i^d)]$. Thus, we have $g_i^P = p - p_R + \alpha p_R \equiv g^P(p, p_R)$ for each $i \in P$. Finally, by using $v(g) = g$, $c(g) = \frac{g^2}{2}$, g_k^d $(k = F, R)$ and g^P, $\mathcal{S}(P)$ is transformed to

$$S(p, p_R) \equiv \frac{p_R\left((p-2)p_R + 1\right)\alpha^2}{2} + \frac{2(p - p_R)(p-2)p_R\alpha}{2} + \frac{(p - p_R)\left((p-1)^2 + (2-p)p_R\right)}{2}.$$

By the property of the NBS, the transfers that $i \in P$ receives are $T_i^P = -[\alpha_i v(\sum_{j\in P} g^P) - c(g^P)] + \alpha_i v(\sum_{j\in P} g_j^d) - c(g_i^d) + \frac{S(p,p_R)}{p} = -\left[\alpha_i p g^P(p, p_R) - \frac{\left(g^P(p,p_R)\right)^2}{2}\right] + \alpha_i\left((p - p_R)g_F^d + p_R g_F^d\right) - \frac{\left(g_i^d\right)^2}{2} + \frac{S(p,p_R)}{p}$.

The following proposition is the main result of this analysis.

Proposition 1 *There always exists a SPNE at which all FGs choose action F in Stage 1.*

The proof is provided in the appendix. Proposition 1 shows that no countries delegate in a SPNE if participation in a negotiation is voluntary. This is quite different from the results when the participation decision is omitted from the analysis. Eckert (2003) shows that in the case of $n = 2$, without the participation decision, (i) every FG delegates negotiation to the polluter region if the region is sufficiently populous and (ii) it is never supported in any SPNE where all FGs choose nondelegation. These results can be generalized to $n \geq 3$ identical countries.[5]

The delegation has both advantages and disadvantages for the FG. The disadvantage arises from the failure to internalize the entire country's preferences because the RG makes the decisions. The benefit of delegation is the strategic effect that it has in improving the FG's bargaining position. If the FG of a country delegates the negotiation to the polluter RG, then the FG can manipulate the negotiation outcome because it will reflect the RG's preferences for less pollution abatement. This will benefit the FG using delegation for this purpose. In contrast, given that some FGs use delegation in this way, the manipulated bargaining outcome is disadvantageous to the FG that chooses

[3] A central characteristic of the voluntary "participation" game in IEAs is that the participation decision of each country depends on the other countries' decisions, and hence no country has a dominant participation action. This is captured under these functions, although each country has a dominant action in the voluntary "contribution" game of a public good. We refer interested readers to Shinohara (2020), the discussion paper version of this study, for the analysis of general functional forms.

[4] See Online Appendix A for a detailed derivation of the Nash bargaining outcome.

[5] We refer readers to Online Appendix D for the generalization.

nondelegation. Without voluntary participation, the benefit exceeds the disadvantage when α is sufficiently high, and hence, every FG chooses delegation.

However, despite the effective manipulation through delegation for high αs, it is interesting that the benefits of delegation are tempered in the case of voluntary participation because the FG that makes the participation decision is unlikely to participate in the negotiation when the manipulation is effective. To obtain this point, we compare the gains of participation between FGs and RGs. Suppose that P with (p, p_R) is a set of participants in Stage 2. A FG outside P obtains the payoff $v((n-n_R)g_F^d + n_R g_R^d) - c(g_F^d) + \frac{S(p+1,p_R)}{p+1}$ if it additionally joins P and the payoff $v\Big(pg^P(p,p_R) + [n-n_R-(p-p_R)]g_F^d + (n_R-p_R)g_R^d\Big) - c(g_F^d)$ if it does not. Then, subtracting the latter payoff from the former yields the gains from participation for the FG:

$$\frac{S(p+1,p_R)}{p+1} - \left[pg^P(p,p_R) - (p-p_R)\,g_F^d - p_R g_R^d\right]. \tag{1}$$

Similarly, the gains from participation for the RG are derived as:

$$\frac{S(p+1,p_R+1)}{p+1} - \alpha\left[pg^P(p,p_R) - (p-p_R)\,g_F^d - p_R g_R^d\right]. \tag{2}$$

Subtracting (2) from (1) and substituting the values of g_i^d, g^P, and $S(p,p_R)$ yields:

$$\frac{(\alpha-1)p}{2(p+1)}\left[2p^2 - 2p - \alpha - 1 + 2\,(1-\alpha-p+\alpha p)\,p_R\right].$$

Thus, the gains for the FG are greater than those for the RG if and only if

$$0 < \alpha < \frac{1}{2p^2-2p-1} \text{ and } \frac{1+\alpha+2p-2p^2}{2-2\alpha-2p+2\alpha p} < p_R \leq p.$$

Because $\frac{1}{2p^2-2p-1}$ takes $\frac{1}{3}$ at $p=2$ and converges to zero as p becomes larger, the gains for the FG exceed those for the RG only for a low α. Thus, the FG that makes the participation decision is unlikely to choose participation if α is high and the manipulation through delegation is effective.[6] Given this effect, if one of the FGs deviates from nondelegation to delegation in Stage 1, then the Nash equilibrium set of participants contracts, which diminishes the benefit of delegation.

4 Extensions

Finally, we discuss two extensions of our basic analysis. First, in the analysis of international environmental problems, it may be important to incorporate the effect of international coordination. Hence, we examine a *Pareto undominated subgame perfect Nash equilibrium* (PU-SPNE), which is a refinement of the SPNE, in Online Appendix E. In the PU-SPNE, in each stage, players are defined as taking actions in a Nash equilibrium that is not Pareto dominated by any other Nash equilibrium. In contrast to Proposition 1, whether all FGs choose F in the PU-SPNE depends on the value of α. However, in the PU-SPNE, all FGs still choose F in Stage 1 if α is within a *sufficiently large* range. In other ranges, some FGs choose R, but the majority of the FGs choose F if n is sufficiently large. Thus, the results in this refined equilibrium are different from those that apply in the absence of the participation decision.

Second, we discuss the effect of two new types of intermediate delegations that fall between actions F and R in the basic model (see Online Appendix F). In one type of delegation, the FG delegates only the decision regarding participation in the negotiation to the RG, whereas in the other,

[6] In Online Appendix B, we show that the FG chooses participation in Nash equilibria only when (1) no RG participates, or the number of RG participants is relatively small or (2) α is sufficiently low, even if the number of RG participants is relatively large. Cases (1) and (2) can be interpreted as those in which the manipulation through delegation does not effectively affect the bargaining outcome (see Remark B.1 in this online appendix).

the FG delegates only the actual negotiation. We show that even if we introduce these new delegation types, we always have a SPNE in which all FGs choose nondelegation. Hence, Proposition 1 is robust to this extension. Similar to action R (referred to as "full delegation" in Online Appendix F), if one FG chooses one of the intermediate delegations and the other FGs choose nondelegation, then the equilibrium number of participants and the level of the public good both decrease. This incentivizes the FG not to deviate from nondelegation.

Appendix: Proof of Proposition 1

Let $P^* \equiv \{1,2,3\} \subseteq N$ and consider the following strategy.

Stage 1. All n FGs choose action F.

Stage 2. Depending on the choice in Stage 1, the participation behavior of each country is defined for all subgames starting from Stage 2 as follows.

(2.1) If all FGs choose action F in Stage 1, then P^* is the set of participants.

(2.2) If the FG of one of the countries (e.g., $i \in N$) chooses action R and the others choose action F in Stage 1, then $P = \{i, j\}$ such that $j \neq i$ is the set of participants.

(2.3) Otherwise, the set of participants is any set supported at a Nash equilibrium.

We show that this strategy is a SPNE. First, we analyze Stage 2.

Claim 1 *P^* is a Nash equilibrium set of participants in Stage 2 after all FGs choose F.*

Proof of Claim 1 All players in this stage are FGs. In this strategy, each $i \in P^*$ obtains the payoff $v(\sum_{j\in P^*} g^P(3,0) + \sum_{j\notin P^*} g_F^d) - c(g^P(3,0)) + T_i^{P^*} = v(ng_F^d) - c(g_F^d) + \frac{S(3,0)}{3} = n + \frac{3}{2}$. If country i deviates from I to O, then its payoff is $v(2g^P(2,0) + (n-2)g_F^d) - c(g_F^d) = n + \frac{3}{2}$. In this strategy, each $j \notin P^*$ obtains the payoff $v(3g^P(3,0) + (n-3)g_F^d) - c(g_F^d) = n + \frac{11}{2}$. If country j deviates from O to I, then it obtains the payoff $v(ng_F^d) - c(g_F^d) + \frac{S(4,0)}{4} = n + 4$. Thus, no country deviates from the above strategy in this stage. ||

Claim 2 *In Stage 2, after the FG of country $i \in N$ chooses R and the others choose F, a two-participant set $P = \{i, j\}$ such that $j \neq i$ is a Nash equilibrium set of participants.*

Proof of Claim 2 i and j are designated as in the statement. In this Stage 2, country i's player is the RG and the other players are the FGs. In this strategy, country i obtains the payoff $\alpha v(g_R^d + (n-1)g_F^d) - c(g_R^d) + \frac{S(2,1)}{2} = \frac{3}{4}\alpha^2 + (n-1)\alpha + \frac{1}{4}$. If country i deviates from I to O, then its payoff is $\alpha v(g_R^d + (n-1)g_F^d) - c(g_R^d) = -\frac{1}{2}\alpha^2 + (n-1)\alpha$. Clearly, the former is greater than the latter. In the above strategy, country j obtains the payoff $v(g_R^d + (n-1)g_F^d) - c(g_F^d) + \frac{S(2,1)}{2} = \frac{1}{4}\alpha^2 + \alpha + n - \frac{3}{4}$. If j deviates from I to O, then its payoff is $v(g_R^d + (n-1)g_F^d) - c(g_F^d) = \alpha + n - \frac{3}{2}$. Clearly, the former is greater than the latter. Finally, in the above strategy, country $k \notin P$ obtains the payoff $v(2g^P(2,1) + (n-2)g_F^d) - c(g_F^d) = 2\alpha + n - \frac{1}{2}$. If k deviates from O to I, then its payoff is $v(g_R^d + (n-1)g_F^d) - c(g_F^d) + \frac{S(3,1)}{3} = \frac{1}{3}\alpha^2 + \frac{5}{3}\alpha + n - \frac{1}{2}$. By some calculation, the former is greater than the latter. Thus, no country deviates from the above strategy in this stage. ||

Finally, we analyze Stage 1. In the above strategy, the FG of country $i \in P^*$ obtains the payoff $v(\sum_{j\in P^*} g^P(3,0) + \sum_{j\notin P^*} g_F^d) - c(g^P(3,0)) + T_i^{P^*} = v(ng_F^d) - c(g_F^d) + \frac{S(3,0)}{3} = n + \frac{3}{2}$ and the FG of country $i \notin P^*$ obtains the payoff $v(\sum_{j\in P^*} g^P(3,0) + \sum_{j\notin P^*} g_F^d) - c(g_F^d) = n + \frac{11}{2}$. If one of

the countries in P^* (e.g., country 1) deviates from F to R, then $P = \{1, j\}$, such that $j \neq 1$ is the equilibrium set of participants in Stage 2 after this deviation by Claim 2. Then, each country in $P = \{1, j\}$ produces $g^P(2,1) = 1 + \alpha$ and country 1's payoff is:

$$\begin{aligned} & v\Big(\sum_{j\in P} g^P(2,1) + \sum_{j\notin P} g_F^d\Big) - c(g^P(2,1)) + T_i^P \\ &= (1-\alpha)\Big(\sum_{j\in P} g^P(2,1) + \sum_{j\notin P} g_F^d\Big) + \alpha\Big(\sum_{j\in P} g_j^d + \sum_{j\notin P} g_F^d\Big) - c(g_R^d) + \frac{S(2,1)}{2} \qquad (3) \\ &= (1-\alpha)\left(2g^P(2,1)\right) + \alpha\Big(g_R^d + g_F^d\Big) + (n-2)g_F^d - \frac{\left(g_R^d\right)^2}{2} + \frac{S(2,1)}{2} = -\frac{5}{4}\alpha^2 + \alpha + \frac{1}{4} + n. \end{aligned}$$

If one of the countries $i \notin P^*$ deviates from F to R, then $P = \{i, j\}$ such that $j \neq i$ is the equilibrium set of participants in Stage 2 by Claim 2, and country i receives the same payoff as (3). By some calculation, we have $n + \frac{11}{2} > n + \frac{3}{2} > -\frac{5}{4}\alpha^2 + \alpha + \frac{1}{4} + n$ if $0 < \alpha < 1$. Hence, no country deviates from the strategy above in Stage 1. In conclusion, it is a SPNE. ■

References

[1] Barrett, S., 1994. Self-enforcing international environmental agreements. *Oxford Economic Papers* 46, 878–894.

[2] Barrett, S., 2006. Climate treaties and "breakthrough" technologies. *American Economic Review* 96, 22–25.

[3] Besley, T., Coate, S., 2003. Centralized versus decentralized provision of local public goods. *Journal of Public Economics* 87, 2611–2637.

[4] Buchholz, W., Haupt, A., Peters, W., 2013. International environmental agreements, fiscal federalism, and constitutional design. *Review of International Economics* 21(4), 705–718.

[5] Carraro, C., Siniscalco, D., 1993. Strategies for the international protection of the environment. *Journal of Public Economics* 52(3), 309–328.

[6] Dalmazzone, S., 2006. Decentralization and the environment. In Ahmad, E., Brosio, G. eds. Handbook of Fiscal Federalism, Edward Elgar, 459–477.

[7] Eckert, H., 2003. Negotiating environmental agreements: Regional or federal authority? *Journal of Environmental Economics and Management* 46, 1–24.

[8] Finus, M., Pintassilgo, P., 2012. International environmental agreements under uncertainty: Does the "veil of uncertainty" help? *Oxford Economic Papers* 64, 736–764.

[9] Goeschl, T., Perino, G., 2017. The climate policy hold-up: Green technologies, intellectual property rights, and the abatement incentives of international agreements. *Scandinavian Journal of Economics* 119(3), 709–732.

[10] Rubio, S.J., Ulph, A., 2006. Self-enforcing international environmental agreements revisited. *Oxford Economic Papers* 58(2), 233–263.

[11] Segendorff, B., 1998. Delegation and threat in bargaining. *Games and Economic Behavior* 23, 266–283.

[12] Shinohara, R., 2020. Voluntary participation in international environmental agreements and authority structures in a federation. Available at SSRN: http://dx.doi.org/10.2139/ssrn.3513664.

Online Appendix

In this online appendix, we present the derivation of the Nash bargaining outcome and the second-stage payoff in Section A. In Section B, we characterize the Nash equilibrium set of participants in Stage 2 for an α that is not too low. In Section C, we calculate the payoffs to the first-stage players (that is, the FGs). In Section D, we present the analysis when the participation decision is absent. In Section E, we present the analysis of the refined SPNE. Finally, we extend our analysis to the case of partial delegation in Section F. The analyses in Sections B and C are helpful steps to examine the refinement of the SPNE in Section E.

A Nash bargaining outcome and payoffs

A.1 Detailed derivation of the Nash bargaining outcome in Section 3

We introduce a parameter α_i such that $\alpha_i = 1$ if country i's second-stage decision maker is the FG and $\alpha_i = \alpha$ if it is the RG of the polluter region. For a set of participants $P \subseteq N$, the set of participating RGs is denoted by $P_R \equiv \{i \in P | \alpha_i = \alpha\}$. The set of participating FGs is automatically determined by $P \setminus P_R = \{i \in P | \alpha_i = 1\}$.

First, we calculate the breakdown outcome of the negotiation. If the negotiation breaks down, then the negotiator of country $i \in P$ determines the abatement level to maximize its payoff. There are no transfers in this case. If we denote country i's public good level at the time of the disagreement by g_i^d, then $g_i^d = \alpha_i$. By this, when the negotiation breaks down, every participant in P_R ($P \setminus P_R$) selects the same abatement level, which is denoted by $g_R^d \equiv \alpha$ ($g_F^d \equiv 1$, respectively). Similarly, every nonparticipant $j \notin P$ decides on the level g_j^d, irrespective of whether the negotiation succeeds. Note that every nonparticipant $j \notin P$ produces g_j^d, irrespective of whether the negotiation succeeds.

In the negotiation, the participants negotiate $(g_j)_{j\in P}$ and $(T_j)_{j\in P}$. To ensure correspondence between the negotiation outcomes and set of participants, we denote the levels of pollution reduction and those of transfers when P is the set of participants by $(g_j^P)_{j\in P}$ and $(T_j^P)_{j\in P}$, respectively. By the negotiation, the negotiator of participating country i receives the payoff:

$$\alpha_i v\Big(\sum_{j\in P} g_j^P + \sum_{j\in N\setminus P} g_j^d\Big) - c(g_i^P) + T_i^P. \tag{4}$$

The NBS maximizes the participants' surplus relative to the breakdown of the negotiation, as follows: $\mathcal{S}(P) \equiv \sum_{i\in P}[\alpha_i v(\sum_{j\in P} g_j^P + \sum_{j\in N\setminus P} g_j^d) - c(g_i^P)] - \sum_{i\in P}[\alpha_i v(\sum_{j\in N} g_j^d) - c(g_i^d)]$. The first summation is the sum of the participants' surplus when the negotiation succeeds and the second summation is the sum of the participants' surplus when the negotiation breaks down. As the second summation is taken as fixed in the negotiation, the negotiation achieves $(g_j^P)_{j\in P}$ to maximize the first summation. From the differentiation of $\mathcal{S}(P)$ with respect to g_i^P and the first-order condition, if the negotiation succeeds, then every participant i invests at the same level of $g^P(p, p_R)$, such that $g^P(p, p_R) \equiv g_i^P = p - p_R + \alpha p_R$. Sometimes, we denote $g^P(p, p_R)$ by g^P for brevity.

Based on the outcome above, $S(P)$ is calculated such that it depends on p and p_R as follows:

$$\begin{aligned}
S(P) =& \sum_{i\in P}\left(\alpha_i v\Big(\sum_{j\in P} g_j^P\Big) - c(g_i^P)\right) - \sum_{i\in P}\left(\alpha_i v\Big(\sum_{j\in P} g_j^d\Big) - c(g_i^d)\right)\\
=& \sum_{i\in P\setminus P_R}\left(v\Big(\sum_{j\in P} g_j^P\Big) - c(g_i^P)\right) + \sum_{i\in P_R}\left(\alpha v\Big(\sum_{j\in P} g_j^P\Big) - c(g_i^P)\right)\\
&- \sum_{i\in P\setminus P_R}\left(v\Big(\sum_{j\in P} g_j^d\Big) - c(g_i^d)\right) - \sum_{i\in P_R}\left(\alpha v\Big(\sum_{j\in P} g_j^d\Big) - c(g_i^d)\right)\\
=& (p-p_R)\left(pg^P - \frac{(g^P)^2}{2}\right) + p_R\left(p\alpha g^P - \frac{(g^P)^2}{2}\right)\\
&- (p-p_R)\left((p-p_R)\,g_F^d + p_R g_R^d - \frac{(g_F^d)^2}{2}\right) - p_R\left(\alpha\left((p-p_R)\,g_F^d + p_R g_R^d\right) - \frac{(g_R^d)^2}{2}\right).
\end{aligned}$$

Note that $|P\setminus P_R| = p - p_R$. Finally, substituting g_k^d $(k = F, R)$ and g^P yields:

$$S(p,p_R) = \frac{p_R\left((p-2)p_R+1\right)\alpha^2}{2} + \frac{2(p-p_R)(p-2)p_R\alpha}{2} + \frac{(p-p_R)\left((p-1)^2+(2-p)p_R\right)}{2}.$$

The Nash bargaining solution equalizes the net gain of each participant. For all $i, j \in P$, $\alpha_i v(\sum_{k\in P} g_k^P) - c(g_i^P) + T_i - [\alpha_i v(\sum_{k\in P} g_k^d) - c(g_i^d)] = \alpha_j v(\sum_{k\in P} g_k^P) - c(g_j^P) + T_j - [\alpha_j v(\sum_{k\in P} g_k^d) - c(g_j^d)]$. By this property, we have $T_i^P = -[\alpha_i v(\sum_{j\in P} g^P) - c(g^P)] + \alpha_i v(\sum_{j\in P} g_j^d) - c(g_i^d) + \frac{S(p,p_R)}{p}$. By the above notations and the functional forms v and c, we have $T_i^P = -\left[\alpha_i p g^P(p,p_R) - \frac{\left(g^P(p,p_R)\right)^2}{2}\right] + \alpha_i\left((p-p_R)g_F^d + p_R g_F^d\right) - \frac{\left(g_i^d\right)^2}{2} + \frac{S(p,p_R)}{p}$.

A.2 Payoffs to players in Stage 2

Substituting T_i^P into (4) yields the payoff functions of players in Stage 2. As below, they depend on the number of countries choosing R (n_R), the number of participating countries (p), and the number of participating RGs (p_R). The payoff to each FG of the $n - n_R$ countries if it chooses participation (I) is:

$$\begin{aligned}
\pi_F^I(n_R,p,p_R) \equiv& v\left((n-n_R)g_F^d + n_R g_R^d\right) - c(g_F^d) + \frac{S(p,p_R)}{p}\\
=& \frac{p_R((p-2)p_R+1)\alpha^2}{2p} + \frac{((p-p_R)(p-2)p_R + pn_R)\,\alpha}{p}\\
&+ \frac{(p-p_R)\left((p-1)^2+(2-p)p_R\right)}{2p} + n - n_R - \frac{1}{2}
\end{aligned} \tag{5}$$

whereas, if it chooses nonparticipation (O), the payoff is:

$$\begin{aligned}
\pi_F^O(n_R,p,p_R) \equiv& v\left(pg^P(p,p_R) + (n-n_R-(p-p_R))\,g_F^d + (n_R-p_R)\,g_R^d\right) - c(g_F^d)\\
=& (n_R - p_R + pp_R)\alpha + p(p-p_R) + n - n_R - (p-p_R) - \frac{1}{2}.
\end{aligned} \tag{6}$$

Each RG of n_R countries obtains the following payoff if it chooses I:

$$\begin{aligned}\pi_R^I(n_R,p,p_R) \equiv& \alpha v\left((n-n_R)g_F^d + n_R g_R^d\right) - c(g_R^d) + \frac{S(p,p_R)}{p}\\ =& \frac{(p_R((p-2)p_R+1)+2pn_R-p)\alpha^2}{2p} + \frac{((p-2)(p-p_R)p_R+(n-n_R)p)\alpha}{p}\\ &+ \frac{(p-p_R)\left((p-1)^2+(2-p)p_R\right)}{2p}\end{aligned} \tag{7}$$

whereas if it chooses O, the payoff is:

$$\begin{aligned}\pi_R^O(n_R,p,p_R) \equiv& \alpha v\left(pg^P(p,p_R) + (n-n_R-(p-p_R))g_F^d + (n_R-p_R)g_R^d\right) - c(g_R^d)\\ =& \left(n_R+(p-1)p_R-\frac{1}{2}\right)\alpha^2 + (n-n_R+(p-1)(p-p_R))\alpha.\end{aligned} \tag{8}$$

In the above four πs, the subscripts represent the action chosen in Stage 1 and the superscripts represent the participation decision in Stage 2.

The payoff functions (5)–(8) are useful to characterize the Nash equilibrium sets of participants in Section B and to clarify the Pareto dominance relation among the Nash equilibria of Stage 2 in Section E.

B Nash equilibrium sets of participants when α is not too low

B.1 A necessary and sufficient condition for Nash equilibrium sets of participants

If $P \subseteq N$ that consists of p countries and p_R RGs is supported at a Nash equilibrium in Stage 2, then it satisfies *internal stability* (IS) and *external stability* (ES) (D'Aspremont et al., 1983). In our model, IS is equivalent to $\pi_F^I(n_R,p,p_R) \geq \pi_F^O(n_R,p-1,p_R)$ and $\pi_R^I(n_R,p,p_R) \geq \pi_R^O(n_R,p-1,p_R-1)$. The first (second) inequality represents that no participating FG (RG) deviates to O. ES is equivalent to $\pi_F^O(n_R,p,p_R) \geq \pi_F^I(n_R,p+1,p_R)$ and $\pi_R^O(n_R,p,p_R) \geq \pi_R^I(n_R,p+1,p_R+1)$. The first (second) inequality indicates that no nonparticipating FG (RG) deviates to I. By using (5)–(8), we derived IS' and ES' in Lemma B.1 from IS and ES, respectively.

Lemma B.1 *$P \subseteq N$ is a Nash equilibrium set of participants in Stage 2 if and only if:*

$$ISF(p,p_R) \geq 0 \text{ and } ISR(p,p_R) \geq 0, \tag{IS'}$$

$$\text{and } ISF(p+1,p_R) \leq 0 \text{ and } ISR(p+1,p_R+1) \leq 0, \tag{ES'}$$

in which

$$ISF(p,p_R) = \frac{p_R(p_R(p-2)+1)\alpha^2 - 2p_R^2(p-2)\alpha + p_R^2(p-2) - p_R - p(p-3)(p-1)}{2p} \text{ and} \tag{9}$$

$$\begin{aligned}ISR(p,p_R) =& \frac{\left(p_R^2(p-2)+p_R\left(-2p^2+4p+1\right)+2(p-2)p\right)\alpha^2}{2p} - \frac{2(p-2)(p-p_R)^2\alpha}{2p}\\ &+ \frac{(p-p_R)\left((p-1)^2+(2-p)p_R\right)}{2p}.\end{aligned} \tag{10}$$

"ISF" (ISR) is the shorthand for the "internal stability for the FG" (RG).

Lemma B.2 derives the properties of $ISF(p,p_R)$ and $ISR(p,p_R)$, which are useful to characterize Nash equilibrium sets of participants in Stage 2.

Lemma B.2 *(a) $ISF(1,0) = 0$, $ISF(2,0) > 0$, and $ISF(2,1) > 0$.*

(b) $ISF(3,0) = 0$ *and* $ISF(3,1) < 0$. $ISF(3,2) \geq 0$ *if and only if* $0 < \alpha \leq \frac{1}{3}$.

(c) Suppose that $p \geq 4$. *Then,* $ISF(p, p_R) < 0$ *if* $\left[0 \leq p_R \leq p-2\right]$ *or* $\left[p_R = p-1 \text{ and } \frac{6-\sqrt{29}}{7} (\approx 0.0878336) < \alpha < 1\right]$.

(d) $ISR(1,1) = 0$, $ISR(2,1) > 0$, *and* $ISR(2,2) > 0$.

(e) $ISR(3,1) > 0$, $ISR(3,2) > 0$, *and* $ISR(3,3) = 0$.

(f) $ISR(p, p_R) < 0$ *if* $p = p_R \geq 4$. $ISR(4,1) \leq 0$ *if and only if* $6 - \sqrt{29} (\approx 0.614835) \leq \alpha < 1$. $ISR(4,3) \leq 0$ *if and only if* $\frac{\sqrt{37}-2}{11} (\approx 0.37116) \leq \alpha < 1$.

Proof. (a) The proof of (a) is immediate from the property of the NBS.

(b) From (9), we have $ISF(3,0) = 0$, $ISF(3,1) = \frac{(\alpha-1)\alpha}{3} < 0$, and $ISF(3,2) = \frac{(\alpha-1)(3\alpha-1)}{3} \geq 0$ if and only if $0 < \alpha \leq \frac{1}{3}$.

(c) First, suppose that $p \geq 4$ and $0 \leq p_R \leq p-2$. From (9), $ISF(p, p_R) < 0$ if and only if

$$X(\alpha, p_R) \equiv p_R \left(p_R(p-2)+1\right)\alpha^2 - 2p_R^2(p-2)\alpha + p_R^2(p-2) - p_R - p(p-3)(p-1) < 0.$$

The coefficient of α^2, $p_R \left(p_R(p-2)+1\right)$, is positive because $p \geq 4$. Hence, if $X(0, p_R), X(1, p_R) < 0$, then $X(\alpha, p_R) < 0$ for all α such that $0 < \alpha < 1$. By some calculation, $X(1, p_R) = -p(p-1)(p-3) < 0$ because $p \geq 4$ and $X(0, p_R) = (p-2)p_R^2 - p_R - p^3 + 4p^2 - 3p$. The coefficient of p_R^2 in $X(0, p_R)$ is positive. Hence, if $X(0, p_R) < 0$ at $p_R = 0, p-2$, then it is negative at the other values of p_R. By some calculation, $X(0,0) = -p^3 + 4p^2 - 3p = -p(p-3)(p-1) < 0$ and $X(0, p-2) = -2(p^2 - 4p + 3) = -2(p-3)(p-1) < 0$. Finally, we have $X(0, p_R) < 0$.

Second, suppose that $p \geq 4$ and $p_R = p-1$. By some calculation, $X(\alpha, p-1) = \left(p-1\right)\left(\left(p^2 - 3p + 3\right)\alpha^2 + \left(-2p^2 + 6p - 4\right)\alpha + 1\right) < 0$ if and only if

$$Y(p) \equiv \frac{1}{p^2 - 3p + 3}\left(p^2 - 3p + 2 - \sqrt{p^4 - 6p^3 + 12p^2 - 9p + 1}\right) < \alpha < 1.$$

Further, we have

$$\frac{\mathrm{d}Y(p)}{\mathrm{d}p} = \frac{(2p-3)\left(-3p^2 + 9p - 7 + 2\sqrt{p^4 - 6p^3 + 12p^2 - 9p + 1}\right)}{2\left(p^2 - 3p + 3\right)^2 \sqrt{p^4 - 6p^3 + 12p^2 - 9p + 1}}.$$

Because $-3p^2 + 9p - 7 + 2\sqrt{p^4 - 6p^3 + 12p^2 - 9p + 1} < 0$ if $p \geq 4$,[1)] we have $\frac{\mathrm{d}Y(p)}{\mathrm{d}p} < 0$. Finally, we have $Y(4) = \frac{6-\sqrt{29}}{7}$. Therefore, if $p_R = p-1$ and $\frac{6-\sqrt{29}}{7} < \alpha < 1$, then $ISF(p, p-1) < 0$ for all $p \geq 4$.

(d) Immediately from the property of the NBS, $ISR(1,1) = 0$ and $ISR(2, p_R) > 0$ if $1 \leq p_R \leq 2$.

(e) By (10), $ISR(3,1) = \frac{(\alpha-1)(\alpha-3)}{3} > 0$ if $0 < \alpha < 1$, $ISR(3,2) = \frac{1-\alpha}{3} > 0$, and $ISR(3,3) = 0$.

(f) By (10), $ISR(p,p) = -\frac{1}{2}\alpha^2(p-1)(p-3) < 0$ because $p \geq 4$. $ISR(4,1) = \frac{3}{8}\left(\alpha^2 - 12\alpha + 7\right) \leq 0$ if and only if $6 - \sqrt{29} \leq \alpha < 1$. $ISR(4,3) = -\frac{1}{8}\left(11\alpha^2 + 4\alpha - 3\right) \leq 0$ if and only if $\frac{\sqrt{37}-2}{11} \leq \alpha < 1$. ■

[1)] Note that $3p^2 - 9p + 7$, $p^4 - 6p^3 + 12p^2 - 9p + 1 > 0$ if $p \geq 4$. Also, $\left(3p^2 - 9p + 7\right)^2 > 4\left(p^4 - 6p^3 + 12p^2 - 9p + 1\right)^2$ is equivalent with $5\left(p^2 - 3p + 3\right)^2 > 0$.

B.2 Characterization of Nash equilibrium sets of participants

Hereafter, we assume that:

$$\frac{6-\sqrt{29}}{7}\ (\approx 0.0878336)\ <\alpha<1. \tag{11}$$

This range of α *almost* covers the whole range $0<\alpha<1$. Our focus is on such "not too low" αs.

Lemma B.3 *If* (11) *holds, then there is no Nash equilibrium set of participants that consists of more than three countries in Stage 2.*

Proof. We show that in Stage 2, there is no Nash equilibrium set of participants P such that $p\geq 4$. If $p\geq 4$, then by Lemma B.2-(c) and -(f), $ISF(p,0)<0$ and $ISR(p,p)<0$. If P satisfies $1\leq p_R\leq p-1$, then $ISF(p,p_R)<0$ by (11) and Lemma B.2-(c). In conclusion, no P such that $p\geq 4$ satisfies IS. ■

By Lemma B.3, we seek the equilibrium set of participants P such that $0\leq p\leq 3$. It is trivially supported at an equilibrium that no countries participate. Immediately from the property of the NBS, no sets of one participant are supported at any equilibrium. Lemmas B.4 and B.5 clarify which set with two or three participants is supported at a Nash equilibrium when (11) holds.

Lemma B.4 *Suppose that* (11) *holds:*

(a) P with $(p,p_R)=(3,0)$ is a Nash equilibrium set of participants if and only if (i) $n_R=0$ or (ii) $n_R\geq 1$ and $6-\sqrt{29}\leq\alpha<1$.

(b) P with $(p,p_R)=(3,1)$ is not a Nash equilibrium set of participants.

(c) P with $(p,p_R)=(3,2)$ is a Nash equilibrium set of participants if and only if $n_R=2$ and $\frac{6-\sqrt{29}}{7}<\alpha\leq\frac{1}{3}$.

(d) P with $(p,p_R)=(3,3)$ is a Nash equilibrium set of participants.

Proof. (a) By Lemma B.2-(b), IS holds because $ISF(3,0)=0$. By (c) and (f) of Lemma B.2, ES holds because $ISF(4,0)<0$ and $ISR(4,1)\leq 0$ if and only if $n_R\geq 1$ and $6-\sqrt{29}\leq\alpha<1$.

(b) P with $(p,p_R)=(3,1)$ does not satisfy IS because $ISF(3,1)<0$ by Lemma B.2-(b).

(c) When $n_R=2$, P with $(p,p_R)=(3,2)$ is a Nash equilibrium set of participants if and only if $ISF(3,2)\geq 0$, $ISF(4,2)<0$, and $ISR(3,2)>0$ hold. By (b), (c), and (e) of Lemma B.2, $ISF(3,2)\geq 0$ if and only if $0<\alpha\leq\frac{1}{3}$, $ISF(4,2)<0$ when (11) holds, and $ISR(3,2)>0$. Hence, when $n_R=2$, P with $(p,p_R)=(3,2)$ is a Nash equilibrium set of participants if and only if $0<\alpha\leq\frac{1}{3}$. When $n_R\geq 3$, P must additionally satisfy $ISR(4,3)\leq 0$, which is equivalent with $\frac{\sqrt{37}-2}{11}\approx 0.37116\leq\alpha<1$ by Lemma B.2-(f). $ISR(4,2)\leq 0$ is incompatible with $ISF(3,2)\geq 0$. Hence, P is not a Nash equilibrium set of participants when $n_R\geq 3$.

(d) By Lemma B.2-(e), IS holds because $ISR(3,3)=0$. By (c) and (f) of Lemma B.2, ES holds because $ISR(4,4)<0$ and $ISF(4,3)<0$. ■

Lemma B.5 *Suppose that* (11) *holds:*

(a) P with $(p,p_R)=(2,0)$ is a Nash equilibrium set of participants if and only if $n_R=0$.

(b) P with $(p,p_R)=(2,1)$ is a Nash equilibrium set of participants if and only if $n_R=1$.

(c) P with $(p,p_R)=(2,2)$ is a Nash equilibrium set of participants if and only if (i) $n_F=0$ or (ii) $n_F\geq 1$ and $\frac{1}{3}\leq\alpha<1$, where n_F denotes the numver of countries choosing F in Stage 1.

Proof. (a) If $n_R = 0$, then $ISF(2,0) > 0$ and $ISF(3,0) = 0$ are satisfied by (a) and (b) of Lemma B.2. Hence, P with $(p, p_R) = (2,0)$ is a Nash equilibrium set of participants. If $n_R \geq 1$, then P does not satisfy ES because $ISR(3,1) > 0$ by Lemma B.2-(e).

(b) IS holds because $ISF(2,1) > 0$ and $ISR(2,1) > 0$ by (a) and (d) of Lemma B.2. By Lemma B.2-(b), $ISF(3,1) < 0$. Suppose, to the contrary, that $n_R \geq 2$. Then, $ISR(3,2) > 0$ (i.e., ES does not hold) by Lemma B.2-(e). Thus, $n_R = 1$ must be satisfied if P is a Nash equilibrium set of participants.

(c) IS holds because $ISR(2,2) > 0$ by Lemma B.2-(d). ES holds because $ISR(3,3) = 0$ by Lemma B.2-(e) and $ISF(3,2) \leq 0$ if and only if $n_F \geq 1$ and $\frac{1}{3} \leq \alpha < 1$ by Lemma B.2-(b). ■

Remark B.1 From the observation above, we can conclude that the FG is less likely to participate in the negotiation when it is disadvantageous. The cases in which the FG chooses participation in Nash equilibria are summarized as (1) no RG participates or the number of RG participants is relatively small (see (a) of Lemma B.4 and (b) of Lemma B.5) or (2) α is relatively low even if the number of RG participants is relatively large (see (c) of Lemma B.4). Cases (1) and (2) can be interpreted as situations in which manipulation through delegation does not affect the bargaining outcome significantly and, hence, the bargaining disadvantage from nondelegation is relatively small.

C Payoffs to players in Stage 1

Suppose that n_R FGs choose R in Stage 1 and after that, P with (p, p_R) is the set of participants in Stage 2. Substituting the Nash bargaining transfer $T_i^P = -[\alpha_i v(\sum_{j\in P} g^P) - c(g^P)] + \alpha_i v(\sum_{j\in P} g_j^d) - c(g_i^d) + \frac{S(p,p_R)}{p}$ in the payoff functions, we have the payoff functions of players in Stage 1. The FG choosing F in Stage 1 and I in Stage 2 obtains the payoff:

$$\Pi_F^I(n_R, p, p_R) \equiv v\Big(\sum_{j\in P} g^P + \sum_{j\in N\setminus P} g_j^d\Big) - c(g^P) + T_i^P = \pi_F^I(n_R, p, p_R) \text{ in (5)} \tag{12}$$

and the FG choosing F in Stage 1 and O in Stage 2 receives the payoff:

$$\Pi_F^O(n_R, p, p_R) \equiv v\Big(\sum_{j\in P} g^P + \sum_{j\in N\setminus P} g_j^d\Big) - c(g_F^d) = \pi_F^O(n_R, p, p_R) \text{ in (6).} \tag{13}$$

Similarly, the FG choosing R in Stage 1 and O in Stage 2 obtains the payoff:

$$\begin{aligned}
\Pi_R^I(n_R, p, p_R) \equiv& v\Big(\sum_{j\in P} g^P + \sum_{j\notin P} g_j^d\Big) - c(g^P) + T_i^P \\
=& (1-\alpha)\Big(\sum_{j\in P} g^P + \sum_{j\in N\setminus P} g_j^d\Big) + \alpha\Big(\sum_{j\in P} g_j^d + \sum_{j\in N\setminus P} g_j^d\Big) - c(g_R^d) + \frac{S(p,p_R)}{p} \\
=& \frac{[p_R((p-2)p_R+1) - p(2p_R(p-1)+1)]\alpha^2}{2p} \\
&+ \frac{[(p-p_R)(p-2)p_R + p(n_R - (p-1)(p-2p_R))]\alpha}{p} \\
&+ \frac{(p-p_R)\left((p-1)^2 + (2-p)p_R\right)}{2p} + (p-1)(p-p_R) + n - n_R
\end{aligned} \tag{14}$$

and the FG choosing R in Stage 1 and O in Stage 2 receives the payoff:

$$\Pi_R^O(n_R, p, p_R) \equiv v\Big(\sum_{j\in P} g^P + \sum_{j\in N\setminus P} g_j^d\Big) - c(g_R^d) = -\frac{\alpha^2}{2} + (n_R + (p-1)p_R)\alpha + (p-1)(p-p_R) + n - n_R. \tag{15}$$

In the above four Πs, the subscripts represent the action chosen in Stage 1 and the superscripts represent the participation decision in Stage 2. The payoff functions (12)–(15) are useful to clarify the Pareto dominance relation among the Nash equilibria of Stage 1 in Section E.

D A benchmark case: The participation of all countries

We provide the result of the benchmark case, in which the participation of all countries in the negotiation is assumed. In this case, $(p, p_R) = (n, n_R)$ holds. By (12) and (14), the payoff to FGs choosing F is $\Pi_F^I(n_R, n, n_R)$ and that to FGs choosing R is $\Pi_R^I(n_R, n, n_R)$.

Result 1 *Suppose that all the countries participate in a negotiation. Then,* (a) *no SPNE supports that all FGs choose action F.* (b) *There exists a SPNE in which all FGs choose action R if and only if:*

$$\frac{2n-1}{2n^3-4n^2+6n-3} \leq \alpha < 1. \tag{16}$$

(c) *The interval of α in* (16) *becomes larger as the number of countries increases.*

Proof. (a) By (12), the payoff to country i's FG when all FGs choose action F is $\Pi_F^I(0, n, 0) = \frac{n^2}{2}$. By (14), the payoff to country i's FG when it takes action R and the others take action F is

$$\Pi_R^I(1, n, 1) = -\frac{\left(2n^2-2n+1\right)\alpha^2}{2n} - \frac{\left(n^3-4n^2+4n-2\right)\alpha}{n} + \frac{3n^3-6n^2+6n-3}{2n}.$$

We obtain that:

$$\Pi_F^I(0, n, 0) - \Pi_R^I(1, n, 1) = \frac{(\alpha-1)\left(\left(2n^2-2n+1\right)\alpha - \left(-2n^3+6n^2-6n+3\right)\right)}{2n}.$$

By some calculation, we find that $2n^2-2n+1 > 0$ if $n \geq 3$. We also find that $-2n^3+6n^2-6n+3 < 0$ if $n \geq 3$ because $-2n^3+6n^2-6n+3$ is decreasing in n and takes a value of -15 at $n = 3$.[2)] Thus, $\left(2n^2-2n+1\right)\alpha - \left(-2n^3+6n^2-6n+3\right) > 0$ if $n \geq 3$. Finally, by $\alpha < 1$, we find that $\Pi_F^I(0, n, 0) < \Pi_R^I(1, n, 1)$, implying that country i's FG is better off by deviating from actions F to R.

(b) We show that $\Pi_F^I(n-1, n, n-1) \leq \Pi_R^I(n, n, n)$ if and only if (16) holds. By (12) and (14),

$$\Pi_F^I(n-1, n, n-1) = \frac{\left(n^3-4n^2+6n-3\right)\alpha^2}{2n} + \frac{2(n-1)^2\alpha}{n} + \frac{2n-1}{2n} \text{ and } \Pi_R^I(n, n, n) = n^2\alpha - \frac{n^2\alpha^2}{2}.$$

Then:

$$\Pi_F^I(n-1, n, n-1) - \Pi_R^I(n, n, n) = \frac{(\alpha-1)((2n^3-4n^2+6n-3)\alpha - (2n-1))}{2n}.$$

By some calculation, $2n^3-4n^2+6n-3$ is shown to be positive.[3)] Finally, by $0 < \alpha < 1$, $\Pi_F^I(n-1, n, n-1) \leq \Pi_R^I(n, n, n)$ if and only if (16) holds.

(c) We obtain that the left-hand side of (16) converges to 0 as $n \to \infty$ because

$$\frac{2n-1}{2n^3-4n^2+6n-3} = \frac{\frac{2}{n^2} - \frac{1}{n^3}}{2 - \frac{4}{n} + \frac{6}{n^2} - \frac{3}{n^3}} \to 0 \text{ as } n \to \infty. \blacksquare$$

By (a) and (b) of Result 1, while it is impossible for all FGs to negotiate by themselves, all FGs delegate negotiation to their polluter regions if every polluter region is sufficiently populous. See the paragraphs after Proposition 1 in the main text for the intuition of these results. Point (b) is a generalization of the result of Eckert (2003) to n symmetric countries. Point (c) shows that if we interpret the length of the interval in (16) as a measure of the extent to which all FGs choose action R in a SPNE, then the extent is very large when there are many countries.

[2)] Note that $\frac{\mathrm{d}(-2n^3+6n^2-6n+30)}{\mathrm{d}n} = -6(n-1)^2 < 0$.

[3)] Note that $\frac{\mathrm{d}(2n^3 4n^2+6n-3)}{\mathrm{d}n} = 6\left(n-\frac{2}{3}\right)^2 + \frac{10}{3} > 0$, $2n^3-4n^2+6n-3$ is positive at $n = 3$, and $n \geq 3$.

E Pareto undominated subgame perfect Nash equilibria (PU-SPNE)

We examine a PU-SPNE of the basic model . In this equilibrium, players take action in a Nash equilibrium that is not Pareto dominated by any other Nash equilibrium in each stage. Thus, in Stage 2 , we derive that the set of participants is supported at a Nash equilibrium that is not Pareto dominated by any other Nash equilibrium. Then, given the equilibrium outcome of Stage 2, in Stage 1, we derive the delegation decisions in a Nash equilibrium that is not Pareto dominated by any other Nash equilibrium.

E.1 Pareto undominated Nash equilibrium sets of participants

First, we derive the set of participants that is supported at a Pareto undominated Nash equilibrium (PUNE) for each second-stage subgame. Henceforth, each second-stage game is characterized by numbers n_F (the number of countries choosing F in Stage 1) and n_R (the number of countries choosing R in Stage 1).

Lemma E.1 lists the set of participants in the second stage in PUNE.[4)]

Lemma E.1 *(a) Consider second-stage subgames with $n_F \geq 3$ and $n_R \geq 3$. If $\frac{6-\sqrt{29}}{7} < \alpha < 6-\sqrt{29}$, then P with $(p, p_R) = (3,3)$ is the only set of participants at a PUNE. If $6-\sqrt{29} \leq \alpha < 1$, then P with $(p, p_R) = (3,0)$ and P with $(p, p_R) = (3,3)$ are the sets of participants at PUNE.*

(b) Consider second-stage subgames with $n_F \geq 3$ and $n_R = 2$. If $\frac{6-\sqrt{29}}{7} < \alpha < \frac{1}{3}$, then P with $(p, p_R) = (3,2)$ is the only set of participants at a PUNE. If $\frac{1}{3} \leq \alpha < 6-\sqrt{29}$, then P with $(p, p_R) = (2,2)$ is the only set of participants at a PUNE. If $6-\sqrt{29} \leq \alpha < 1$, then P with $(p, p_R) = (3,0)$ is the only set of participants at a PUNE.

(c) Consider second-stage subgames with $n_F \geq 3$ and $n_R = 1$. If $\frac{6-\sqrt{29}}{7} < \alpha < 6-\sqrt{29}$, then P with $(p, p_R) = (2,1)$ is the only set of participants at a PUNE. If $6-\sqrt{29} \leq \alpha < 1$, then P with $(p, p_R) = (3,0)$ is the only set of participants at a PUNE.

(d) Consider second-stage subgames such that $n_F \geq 3$ and $n_R = 0$. P with $(p, p_R) = (3,0)$ is the only set of participants at a PUNE.

(e) Consider second-stage subgames such that $0 \leq n_F \leq 2$ and $n_R \geq 3$. P with $(p, p_R) = (3,3)$ is the only set of participants at a PUNE.

E.2 The analysis of Stage 1

Our analysis is focused on the case in which the number of countries is sufficiently large such that $n \geq 5$. Later, we briefly introduce the results in the cases of $n = 3$ and 4.

Result 2 is the main result of appendix E.

Result 2 *Suppose that $n \geq 5$. Then, all FGs choose F in Stage 1 (i.e., $n_R = 0$) in every* PU-SPNE *if $\frac{6-\sqrt{29}}{7} < \alpha < 6-\sqrt{29}$. Three FGs choose R and $n-3$ FGs choose F in Stage 1 (i.e., $n_R = 3$) in every* PU-SPNE *if $6-\sqrt{29} \leq \alpha < 1$.*[5)]

An explanation of the implications of Result 2 is in order. First, we learn from this result that no PU-SPNE supports that all FGs choose R in Stage 1. This is quite different from the case in the absence of the voluntary participation decision, in which it is likely that there is a SPNE in which

[4)] The proof is relegated to Section E.3.

[5)] The proof is relegated to Section E.3.

all FGs choose R. Second, although there may be multiple subgame perfect Nash equilibria in the game with the voluntary participation decision, the SPNE in Proposition 1 in the main text survives the refinement based on the Pareto dominance relation of the equilibria in some sufficiently large range of α ($\frac{6-\sqrt{29}}{7} < \alpha < 6-\sqrt{29}$). In the other range of α ($6-\sqrt{29} \leq \alpha < 1$), three FGs choose R, but the other FGs choose F. Even in this range, if n is sufficiently large, many FGs choose F in the equilibrium. As shown in Result 3, no SPNE supports the choice of R by three FGs in the benchmark model in appendix D.

Result 3 *In the absence of the voluntary participation decision, it is never supported at any SPNE that three FGs choose R and $n-3$ FGs choose F (i.e., $n_R = 3$) if $n \geq 5$.*[6)]

In comparison with Results 1–3, we conclude that completely different delegation structures are observed in the presence and absence of the voluntary participation decision even when we consider a refined SPNE.

Remark E.1 Our analysis in appendix E is limited to the case of $n \geq 5$. We briefly introduce the result in the case of $n = 3$ or 4. In the case of $n = 3$, we can prove that all FGs choose F in Stage 1 (i.e., $n_R = 0$) in every PU-SPNE. In the case of $n = 4$, n_R attained at the PU-SPNE is zero if $\frac{6-\sqrt{29}}{7} < \alpha < 6-\sqrt{29}$, 1 if $6-\sqrt{29} \leq \alpha \leq -14+\sqrt{217}$, 1 or 3 if $-14+\sqrt{217} < \alpha \leq \frac{7}{9}$, and 3 if $\frac{7}{9} < \alpha < 1$. In the whole range of $0 < \alpha < 1$, at least one FG chooses nondelegation, which is similar to Result 2. The proof is available upon request.

E.3 Proofs

Proof of Lemma E.1

(a) By Lemmas B.3–B.5:

- P with $(p, p_R) = (3, 0)$ is a Nash equilibrium set of participants if and only if $6-\sqrt{29} \leq \alpha < 1$.
- P with $(p, p_R) = (3, 3)$ is a Nash equilibrium set of participants if and only if $\frac{6-\sqrt{29}}{7} < \alpha < 1$.
- P with $(p, p_R) = (2, 2)$ is a Nash equilibrium set of participants if and only if $\frac{1}{3} \leq \alpha < 1$.

We examine the Pareto dominance relation among these equilibria. If $6-\sqrt{29} \leq \alpha < 1$, then the Nash equilibria with $(p, p_R) = (3, 0)$ and $(p, p_R) = (3, 3)$ coexist in the game of Stage 2. First, we show that these two equilibria do not have any Pareto dominance relation. A country that chooses participation in the equilibrium with $(p, p_R) = (3, 0)$ obtains the payoff $\pi_F^I(n_R, 3, 0) = \frac{3}{2} + n - n_R + \alpha n_R$. This country does not participate in the equilibrium with $(p, p_R) = (3, 3)$ and, hence, it obtains the payoff $\pi_F^O(n_R, 3, 3) = -\frac{1}{2} + n - n_R + \alpha(6 + n_R)$ in the equilibrium. We have $\pi_F^I(n_R, 3, 0) < \pi_F^O(n_R, 3, 3)$ if $6-\sqrt{29} \leq \alpha < 1$. On the other hand, there is a RG that does not participate in the equilibrium with $(p, p_R) = (3, 0)$, but does participate in the equilibrium with $(p, p_R) = (3, 3)$. It obtains the payoff $\pi_R^O(n_R, 3, 0) = \alpha(6 + n - n_R) + \alpha^2\left(-\frac{1}{2} + n_R\right)$ in the first equilibrium and $\pi_R^I(n_R, 3, 3) = \alpha(n - n_R) + \alpha^2\left(\frac{3}{2} + n_R\right)$ in the second. By some calculation, we have $\pi_R^I(n_R, 3, 3) < \pi_R^O(n_R, 3, 0)$. Thus, the Nash equilibria do not have any Pareto dominance relation if $6-\sqrt{29} \leq \alpha < 1$.

Second, we show that there is a Nash equilibrium with $(p, p_R) = (3, 3)$ that Pareto dominates the equilibrium with $(p, p_R) = (2, 2)$. Let $P = \{j, k\}$ be the set of participants in the equilibrium with $(p, p_R) = (2, 2)$. In addition, suppose that $Q = \{j, k, \ell\}$ is the set of participants in the equilibrium

[6)] The proof is relegated to Section E.3.

with $(p, p_R) = (3, 3)$ such that $\ell \notin P$ (note that j and k are participants in both equilibria). Each of j and k obtains the payoff $\pi_R^I(n_R, 3, 3)$ in the equilibrium with $(p, p_R) = (3, 3)$ and $\pi_R^I(n_R, 2, 2)$ in that with $(p, p_R) = (2, 2)$. By some calculation, we have $\pi_R^I(n_R, 3, 3) = \alpha(n - n_R) + \alpha^2\left(\frac{3}{2} + n_R\right) > \alpha(n - n_R) + \alpha^2 n_R = \pi_R^I(n_R, 2, 2)$. Country ℓ obtains the payoff $\pi_R^I(n_R, 3, 3)$ in the equilibrium with $(p, p_R) = (3, 3)$ and $\pi_R^O(n_R, 2, 2)$ in the equilibrium with $(p, p_R) = (2, 2)$. By some calculation, we have $\pi_R^I(n_R, 3, 3) = \alpha(n - n_R) + \alpha^2\left(\frac{3}{2} + n_R\right) = \pi_R^O(n_R, 2, 2)$. Finally, for the other countries, we have $\pi_F^O(n_R, 3, 3) = -\frac{1}{2} + n - n_R + \alpha(6 + n_R) > -\frac{1}{2} + n - n_R + \alpha(2 + n_R) = \pi_F^O(n_R, 2, 2)$ and $\pi_R^O(n_R, 3, 3) = \alpha^2\left(n_R + \frac{11}{2}\right) + \alpha(n - n_R) > \alpha^2\left(n_R + \frac{3}{2}\right) + \alpha(n - n_R) = \pi_R^O(n_R, 2, 2)$. These inequalities show that the Nash equilibrium with $(p, p_R) = (2, 2)$ is Pareto dominated by the equilibrium with $(p, p_R) = (3, 3)$.

(b) By Lemmas B.3–B.5, in this subgame, we have that:

- P with $(p, p_R) = (3, 0)$ is a Nash equilibrium set of participants if and only if $6 - \sqrt{29} \leq \alpha < 1$.
- P with $(p, p_R) = (3, 2)$ is a Nash equilibrium set of participants if and only if $\frac{6-\sqrt{29}}{7} < \alpha \leq \frac{1}{3}$.
- P with $(p, p_R) = (2, 2)$ is a Nash equilibrium set of participants if and only if $\frac{1}{3} \leq \alpha < 1$.

If $6 - \sqrt{29} \leq \alpha < 1$, then the Nash equilibrium with $(p, p_R) = (3, 0)$ and that with $(p, p_R) = (2, 2)$ coexist. Let P (Q) be the set of participants in the equilibria with $(p, p_R) = (3, 0)$ ($(p, p_R) = (2, 2)$). Note that P and Q are disjointed. For every country in $P \setminus Q$, $\pi_F^I(n_R, 3, 0) = \frac{3}{2} + n - n_R + \alpha n_R > \pi_F^O(n_R, 2, 2) = -\frac{1}{2} + n - n_R + \alpha(2 + n_R)$. For every country in $Q \setminus P$, $\pi_R^O(n_R, 3, 0) = \alpha(6 + n - n_R) + \alpha^2\left(-\frac{1}{2} + n_R\right) > \pi_R^I(n_R, 2, 2) = \alpha(n - n_R) + \alpha^2 n_R$. For every country in $(N \setminus P) \cap (N \setminus Q)$, $\pi_F^O(n_R, 3, 0) = \alpha n_R + n - n_R + \frac{11}{2} > \alpha(n - n_R) + n - n_R - \frac{1}{2} = \pi_F^O(n_R, 2, 2)$ and $\pi_R^O(n_R, 3, 0) = \left(n_R - \frac{1}{2}\right)\alpha^2 + (n - n_R + 6)\alpha > \left(n_R + \frac{3}{2}\right)\alpha^2 + (n - n_R)\alpha = \pi_R^O(n_R, 2, 2)$. By these conditions, the Nash equilibrium with $(p, p_R) = (3, 0)$ Pareto dominates that with $(p, p_R) = (2, 2)$.

(c) By Lemmas B.3–B.5, in this subgame, we have that:

- P with $(p, p_R) = (3, 0)$ is a Nash equilibrium set of participants if and only if $6 - \sqrt{29} \leq \alpha < 1$.
- P with $(p, p_R) = (2, 1)$ is a Nash equilibrium set of participants.

If $6 - \sqrt{29} \leq \alpha < 1$, then the Nash equilibrium with $(p, p_R) = (3, 0)$ and that with $(p, p_R) = (2, 1)$ coexist. Let $Q = \{j, k\}$ be the set of participants in the Nash equilibrium with $(p, p_R) = (2, 1)$, in which the participation decision of country j (k) is made by the FG (RG). In the Nash equilibrium with $(p, p_R) = (3, 0)$, let $P = \{j, \ell, m\}$ be the set of participants. The participation decision of every country in P is made by the FG. We show that the equilibrium with the set of participants Q is Pareto dominated by that with P. For the country in $P \cap Q = \{j\}$, $\pi_F^I(n_R, 3, 0) = \frac{3}{2} + n - n_R + \alpha n_R > \pi_F^I(n_R, 2, 1) = \frac{\alpha^2}{4} + \alpha n_R + n - n_R - \frac{1}{4}$. For every country in $P \setminus Q = \{\ell, m\}$, $\pi_F^I(n_R, 3, 0) = \frac{3}{2} + n - n_R + \alpha n_R > \pi_F^O(n_R, 2, 1) = \frac{1}{2} + n - n_R + \alpha(1 + n_R)$. For every country in $Q \setminus P = \{k\}$, $\pi_R^O(n_R, 3, 0) = \alpha^2\left(-\frac{1}{2} + n_R\right) + \alpha(6 + n - n_R) > \pi_R^I(n_R, 2, 1) = \alpha^2\left(n_R - \frac{1}{4}\right) + \alpha(n - n_R) + \frac{1}{4}$ if $6 - \sqrt{29} \leq \alpha < 1$. For every country in $(N \setminus P) \cap (N \setminus Q)$, $\pi_F^O(n_R, 3, 0) = \alpha n_R + n - n_R + \frac{11}{2} > \pi_F^O(n_R, 2, 1) = \alpha(n_R + 1) + n - n_R + \frac{1}{2}$. By these conditions, the Nash equilibrium with $(p, p_R) = (3, 0)$ Pareto dominates that with $(p, p_R) = (2, 1)$.

(d) By Lemmas B.3–B.5, in this subgame, we have that:

- P with $(p, p_R) = (3, 0)$ is a Nash equilibrium set of participants.

- P with $(p, p_R) = (2, 0)$ is a Nash equilibrium set of participants.

The Nash equilibrium with $(p, p_R) = (3, 0)$ Pareto dominates that with $(p, p_R) = (2, 0)$ because $\pi_F^I(0, 3, 0) = n + 4 > n + \frac{3}{2} = \pi_F^I(0, 2, 0)$, $\pi_F^I(0, 3, 0) = n + \frac{3}{2} = \pi_F^O(0, 2, 0)$, $\pi_F^O(0, 3, 0) = n + \frac{11}{2} > n + \frac{3}{2} = \pi_F^I(0, 2, 0)$, and $\pi_F^O(0, 3, 0) = n + \frac{11}{2} > n + \frac{3}{2} = \pi_F^O(0, 2, 0)$.

(e) By Lemmas B.3–B.5, in this subgame, we have that:

- P with $(p, p_R) = (3, 3)$ is a Nash equilibrium set of participants.
- P with $(p, p_R) = (2, 2)$ is a Nash equilibrium set of participants if $[1 \leq n_F \leq 2$ and $\frac{1}{3} \leq \alpha < 1]$ or $n_F = 0$.

We can show that the Nash equilibrium with $(p, p_R) = (3, 3)$ Pareto dominates that with $(p, p_R) = (2, 2)$ in a similar way to (a) in this proof. ■

Proof of Result 2

Case 1. $\frac{6-\sqrt{29}}{7} < \alpha < 6 - \sqrt{29}$

In this case, given the second-stage equilibrium in Lemma E.1, we show that (A) there is a SPNE that supports $n_R = 0$ and (B) $n_R \geq 1$ is never supported by any PU-SPNE.

(A) In Stage 2, after all FGs choose F in Stage 1, P with $(p, p_R) = (3, 0)$ is the set of participants at any PUNE by Lemma E.1-(d). Take $i \in P$. Given the other FGs' choice, country i's FG receives the payoff $\Pi_F^I(0, 3, 0) = n + \frac{3}{2}$ if it chooses F in Stage 1. If country i's FG deviates to R, then P' satisfying $(p', p_R') = (2, 1)$ is the set of participants supported at any PUNE by Lemma E.1-(c). After this deviation, country i is the only country for which the RG makes the participation decision. Hence, $i \in P'$ and the payoff is $\Pi_R^I(1, 2, 1) = -\frac{5}{4}\alpha^2 + \alpha + n + \frac{1}{4}$. By some calculation, we have $\Pi_F^I(0, 3, 0) > \Pi_R^I(1, 2, 1)$. Take $j \notin P$. By similar reasoning, given the other countries' choice in Stage 1, country j's FG receives the payoff $\Pi_F^O(0, 3, 0) = n + \frac{11}{2}$ if it chooses F and $\Pi_R^I(1, 2, 1) = -\frac{5}{4}\alpha^2 + \alpha + n + \frac{1}{4}$. By some calculation, we have $\Pi_F^O(0, 3, 0) > \Pi_R^I(1, 2, 1)$. Thus, no FG deviates in Stage 1.

(B) In the following claims, we prove that for each integer $n_R \geq 1$ in the first-stage game, given the second-stage outcome in Lemma E.1, there exists no Nash equilibrium supporting that n_R FGs choose R or, if such a Nash equilibrium exists, then it is Pareto dominated by other Nash equilibria.

Claim E.1 *In Case 1, no Nash equilibrium supports that only one country's FG chooses R (i.e., $n_R = 1$) in Stage 1.*

Proof of Claim E.1 When only one country's FG chooses R (i.e., $n_R = 1$), P' satisfying $(p', p_R') = (2, 1)$ is the PUNE set of participants in the subsequent Stage 2 by Lemma E.1-(c). Let i be the country for which the FG chooses R in Stage 1 and $P' = \{i, j\}$ $(i \neq j)$. Then, country i's FG earns the payoff $\Pi_R^I(1, 2, 1) = -\frac{5}{4}\alpha^2 + \alpha + n + \frac{1}{4}$. As we show in part (A), we have $\Pi_F^I(0, 3, 0) > \Pi_R^I(1, 2, 1)$ and $\Pi_F^O(0, 3, 0) > \Pi_R^I(1, 2, 1)$. These conditions prove that country i's FG is made better off by deviating from R to F; that is, $n_R = 1$ is not a Nash equilibrium of the Stage 1 game. ||

Claim E.2 *In Case 1, the first-stage actions in which the FGs of two countries choose R (i.e., $n_R = 2$) is Pareto dominated by the Nash equilibrium with $n_R = 0$.*

Proof of Claim E.2 When the FGs of two countries choose R (i.e., $n_R = 2$), P' satisfying $(p', p_R') = (3, 2)$ if $\frac{6-\sqrt{29}}{7} < \alpha \leq \frac{1}{3}$ and $(p', p_R') = (2, 2)$ if $\frac{1}{3} \leq \alpha < 6 - \sqrt{29}$ is the PUNE set of participants in the subsequent Stage 2 by Lemma E.1-(b).

First, consider the case of $\frac{6-\sqrt{29}}{7} < \alpha \leq \frac{1}{3}$. Let $P' = \{i, j, k\}$ be such that the participation decision is made by the RG for countries i and j and by the FG for country k. Then, the FGs of countries i

and j earn the payoff $\Pi_R^I(2,3,2) = -\frac{7}{2}\alpha^2 + \frac{14}{3}\alpha + n + \frac{1}{3}$ and the FG of country k earns the payoff $\Pi_F^I(2,3,2) = \alpha^2 + \frac{8}{3}\alpha + n - \frac{13}{6}$. The FGs of the other countries earn $\Pi_F^O(2,3,2) = 6\alpha + n - \frac{1}{2}$. Then, suppose that when all countries' FGs choose F (i.e., $n_R = 0$), $P = \{i,j,k\}$ satisfying $(p, p_R) = (3,0)$ is the PUNE set of participants in the subsequent Stage 2. Then, the FGs of all countries in P earn the payoff $\Pi_F^I(0,3,0) = n + \frac{3}{2}$ and those of the other countries earn $\Pi_F^O(0,3,0) = n + \frac{11}{2}$. By some calculation, we have $\Pi_F^I(0,3,0) \geq \Pi_R^I(2,3,2)$ for countries i and j, $\Pi_F^I(0,3,0) \geq \Pi_F^I(2,3,2)$ for country k, and $\Pi_F^O(0,3,0) > \Pi_F^O(2,3,2)$ for the other countries.

Second, consider the case of $\frac{1}{3} \leq \alpha < 6 - \sqrt{29}$. Let $P' = \{i,j\}$ be such that the participation decision is made by the RG for countries i and j. Then, the FGs of countries i and j earn the payoff $\Pi_R^I(2,2,2) = -2\alpha^2 + 4\alpha + n - 2$ and the FGs of the other countries earn the payoff $\Pi_F^O(2,2,2) = 4\alpha + n - \frac{5}{2}$. Then, suppose that when all countries' FGs choose F (i.e., $n_R = 0$), $P = \{i,j,\ell\}$ satisfying $(p, p_R) = (3,0)$ is the PUNE set of participants in the subsequent Stage 2. Then, the FGs of all countries in P earn the payoff $\Pi_F^I(0,3,0) = n + \frac{3}{2}$ and those of the other countries earn $\Pi_F^O(0,3,0) = n + \frac{11}{2}$. By some calculation, we have $\Pi_F^I(0,3,0) > \Pi_R^I(2,2,2)$ for countries i and j, $\Pi_F^I(0,3,0) > \Pi_F^O(2,2,2)$ for country ℓ, and $\Pi_F^O(0,3,0) > \Pi_F^O(2,2,2)$ for the other countries.

Hence, in both cases, the action profile with $n_R = 2$ is Pareto dominated by that with $n_R = 0$. ||

Claim E.3 *In Case 1, the first-stage actions in which the FGs of three countries choose R (i.e., $n_R = 3$) is Pareto dominated by the Nash equilibrium with $n_R = 0$.*

Proof of Claim E.3 We consider the case of $n_R = 3$. When the FGs of three countries choose R (i.e., $n_R = 3$), P' satisfying $(p', p'_R) = (3,3)$ is the PUNE set of participants in the subsequent stage, Stage 2. Assume $P' = \{i,j,k\}$. Then, the FGs of countries in P' earn the payoff $\Pi_R^I(3,3,3) = -\frac{9}{2}\alpha^2 + 9\alpha + n - 3$ and the FGs of the other countries earn the payoff $\Pi_F^O(3,3,3) = 9\alpha + n - \frac{7}{2}$. Then, suppose that when all countries' FGs choose F (i.e., $n_R = 0$), $P = \{i,j,k\}$ satisfying $(p, p_R) = (3,0)$ is the PUNE set of participants in the subsequent Stage 2. Then, the FGs of all countries in P earn the payoff $\Pi_F^I(0,3,0) = n + \frac{3}{2}$ and the FGs of the other countries earn $\Pi_F^O(0,3,0) = n + \frac{11}{2}$. By some calculation, we have $\Pi_F^I(0,3,0) > \Pi_R^I(3,3,3)$ for countries in $P = P'$ and $\Pi_F^O(0,3,0) > \Pi_F^O(3,3,3)$ for the other countries. In Stage 1, the action profile with $n_R = 3$ is Pareto dominated by that with $n_R = 0$. ||

Claim E.4 *In Case 1, no Nash equilibrium supports that the FGs of n_R ($n_R \geq 4$) countries choose R in Stage 1.*

Proof of Claim E.4 Suppose that the FGs of $n_R \geq 4$ countries choose R and those of the remaining countries choose F in Stage 1, and that P is the set of participants at PUNE in the subsequent stage, Stage 2. By Lemma E.1-(a), P satisfies $(p, p_R) = (3,3)$. Take $i \in P$. We obtain from Lemma E.1-(a) that if the FG of country i deviates from R to F in Stage 1 given the choices of the other FGs, then the PUNE set of participants in the subsequent stage, denoted by P', has the same property. That is, P' only consists of countries for which participation decisions are made by the RGs. This implies that country i does not belong to P' after this deviation. In conclusion, the payoff to country i's FG is $\Pi_R^I(n_R,3,3) = -\frac{9}{2}\alpha^2 + \alpha(n_R + 6) + n - n_R$ before the deviation and $\Pi_F^O(n_R - 1,3,3) = \alpha(n_R + 5) + \frac{1}{2} + n - n_R$ after. Then, we have $\Pi_R^I(n_R,3,3) < \Pi_F^O(n_R - 1,3,3)$. Thus, country i's FG is made better off by deviating from R to F in Stage 1. ||

In conclusion, in Case 1, $n_R = 0$ is attained at every PU-SPNE.

Case 2. $6-\sqrt{29} \leq \alpha < 1$

Claim E.5 *In Case 2, given the second-stage outcome in Lemma E.1, n_R is the number of FGs choosing R in a SPNE if and only if $3 \leq n_R \leq n-2$.*

Proof of Claim E.5 First, we show that there exists a SPNE in which n_R ($3 \leq n_R \leq n-2$) FGs choose R. Let N_R be a subset of N such that $|N_R| = n_R$. Consider the following strategy:

In Stage 1, every country in N_R chooses R and every country outside N_R chooses F. In Stage 2, after the set of countries that choose R is N_R, the set of participants is P with $(p, p_R) = (3,3)$. In Stage 2 after the set of countries that choose R is $N_R \setminus \{i\}$ for some $i \in N_R$, the set of participants is P' such that $(p', p'_R) = (3,0)$ and $i \in P'$. In Stage 2, after the set of countries that choose R is $N_R \cup \{i\}$ for some $i \in N \setminus N_R$, the set of participants is P'' such that $(p'', p''_R) = (3,3)$ and $i \in P''$. In any other subgames of Stage 2, the set of participants is any set supported at a Nash equilibrium of the stage game.

In this strategy, the FGs of the countries in $N_R \cap P$, $N_R \setminus P$, and $N \setminus N_R$ obtain the payoffs $\Pi^I_R(n_R,3,3)$, $\Pi^O_R(n_R,3,3)$, and $\Pi^O_F(n_R,3,3)$, respectively. If the FG of country $i \in N_R$ deviates to F in Stage 1, then it obtains the payoff $\Pi^I_F(n_R-1,3,0)$. If the FG of country $i \in N \setminus N_R$ deviates to R in Stage 1, then it obtains the payoff $\Pi^I_R(n_R+1,3,3)$. By some calculation, we have $\Pi^I_R(n_R,3,3) = -\frac{9}{2}\alpha^2 + (n_R+6)\alpha + n - n_R > \Pi^I_F(n_R-1,3,0) = (n_R-1)\alpha + n - n_R + \frac{5}{2}$ for countries in $N_R \cap P$, $\Pi^O_R(n_R,3,3) = -\frac{1}{2}\alpha^2 + (6+n_R)\alpha + n - n_R > \Pi^I_F(n_R-1,3,0) = (n_R-1)\alpha + n - n_R + \frac{5}{2}$ for countries in $N_R \setminus P$, and $\Pi^O_F(n_R,3,3) = (n_R+6)\alpha + n - n_R - \frac{1}{2} > \Pi^I_R(n_R+1,3,3) = -\frac{9}{2}\alpha^2 + (n_R+7)\alpha + n - n_R - 1$ for countries in $N \setminus N_R$. Thus, the strategy is a SPNE.

Second, we prove that there is no SPNE in which n_R FGs choose R in Stage 1 if $0 \leq n_R \leq 2$ or $n-1 \leq n_R \leq n$.

(i) Suppose that $0 \leq n_R \leq 1$. Then, the PUNE set of participants in the subsequent stage is P with $(p, p_R) = (3,0)$ by (c) and (d) of Lemma E.1. Let $i \in P$. If country i's FG switches from F to R given the choice of the other FGs in Stage 1, then P' with $(p', p'_R) = (3,0)$ is the PUNE set of participants in the subsequent stage by (b) and (c) of Lemma E.1. Before the deviation, the payoff to country i's FG is $\Pi^I_F(n_R,3,0) = n_R\alpha + n - n_R + \frac{3}{2}$, whereas after the deviation it is $\Pi^O_R(n_R+1,3,0) = -\frac{1}{2}\alpha^2 + \alpha(1+n_R)n - n_R + 5$. By some calculation, we have $\Pi^I_F(n_R,3,0) < \Pi^O_R(n_R+1,3,0)$. Thus, this case never appears in any Nash equilibrium of Stage 1.

(ii) Suppose that $n_R = 2$. Then, the PUNE set of participants in the subsequent stage is P with $(p, p_R) = (3,0)$ by Lemma E.1-(b). Let $i \in P$. Let P' be the set of participants when country i's FG deviates from F to R given the choice of the other FGs in Stage 1. Then, $(p', p'_R) = (3,3)$ or $(3,0)$ (note that if $n \geq 6$, then $(p', p'_R) = (3,0)$ can be an equilibrium set of participants). The payoff to country i's FG before the deviation is $\Pi^I_F(2,3,0) = 2\alpha + n - \frac{1}{2}$ and that after the deviation is $\Pi^O_R(3,3,0) = -\frac{1}{2}\alpha^2 + 3\alpha + n + 3$ or $\Pi^I_R(3,3,3) = -\frac{9}{2}\alpha^2 + 9\alpha + n - 3$. By some calculation, we have $\Pi^I_F(2,3,0) < \Pi^I_R(3,3,3) < \Pi^O_R(3,3,0)$, which shows that this case never appears in any Nash equilibrium of Stage 1.

(iii) Suppose that $n-1 \leq n_R \leq n$. Then, the PUNE set of participants in the subsequent Stage 2 is P with $(p, p_R) = (3,3)$ by Lemma E.1-(e). Let $i \in P$. If country i's FG switches from R to F given the choice of the other FGs in Stage 1, then P' with $(p', p'_R) = (3,3)$ is the PUNE set of participants. Before the deviation, the payoff to country i's FG is $\Pi^I_R(n_R,3,3) = -\frac{9}{2}\alpha^2 + (n_R+6)\alpha + n - n_R$, whereas after the deviation, it is $\Pi^O_F(n_R-1,3,3) = (n_R+5)\alpha + n - n_R + \frac{1}{2}$. By some calculation, we have $\Pi^I_R(n_R,3,3) < \Pi^O_F(n_R-1,3,3)$. Thus, this case never appears in any Nash equilibrium of Stage 1. ||

Claim E.6 *In Case 2, in every SPNE with* $3 \leq n_R \leq n-2$*, P with* $(p, p_R) = (3,3)$ *is the set of participants in Stage 2 after* n_R *FGs choose R in Stage 1.*

Proof of Claim E.6 Suppose that when n_R FGs choose R in Stage 1, P with $(p, p_R) = (3,0)$ is the set of participants in the subsequent Stage 2. Note that it is supported at a PUNE by Lemma E.1-(a). Let $i \in P$. Country i's FG obtains the payoff $\Pi_F^I(n_R, 3, 0) = n_R\alpha + n - n_R + \frac{3}{2}$. If country i's FG deviates from F to R in Stage 1, then there are two possible equilibrium sets of participants P' with $(p', p'_R) = (3,3)$ and $(p', p'_R) = (3,0)$ (see Lemma E.1-(a)). Country i may belong to P' with $(p', p'_R) = (3,3)$. By this deviation, country i's FG earns $\Pi_R^I(n_R+1,3,3) = -\frac{9}{2}\alpha^2 + (n_R+7)\alpha + n - n_R - 1$, $\Pi_R^O(n_R+1,3,3) = -\frac{1}{2}\alpha^2 + (n_R+7)\alpha + n - n_R - 1$, or $\Pi_R^O(n_R+1,3,0) = -\frac{1}{2}\alpha^2 + (n_R+1)\alpha + n - n_R - 5$. By some calculation, we have $\Pi_F^I(n_R,3,0) < \Pi_R^I(n_R+1,3,3)$, $\Pi_F^I(n_R,3,0) < \Pi_R^O(n_R+1,3,3)$, and $\Pi_F^I(n_R,3,0) < \Pi_R^O(n_R+1,3,0)$ by $6 - \sqrt{29} \leq \alpha < 1$. Hence, country i's FG is made better off by switching from F to R for any anticipation of the second-stage equilibrium. In conclusion, there is no SPNE with $3 \leq n_R \leq n-2$ that supports P with $(p, p_R) = (3,0)$ as the set of participants in Stage 2. ||

Claim E.7 *In Case 2, no SPNE with* $4 \leq n_R \leq n-2$ *is a PU-SPNE.*

Proof of Claim E.7 Let $\tilde{n}_R$ be the number such that $4 \leq \tilde{n}_R \leq n-2$. Take a SPNE in which $\tilde{n}_R$ FGs choose R in Stage 1 and $\tilde{P}$ is the PUNE set of participants in Stage 2. Then, $\tilde{P}$ satisfies $(\tilde{p}, \tilde{p}_R) = (3,3)$ (see Claim E.6). In addition, there exists another SPNE with $n_R = 3$ such that the same set $\tilde{P}$ with $(\tilde{p}, \tilde{p}_R) = (3,3)$ is the PUNE set of participants.

We show that in Stage 1, given PUNE sets of participants for every Stage 2 subgame, the SPNE with $\tilde{n}_R$ is Pareto dominated by that with $n_R = 3$. We have $\Pi_R^I(3,3,3) = -\frac{9}{2}\alpha^2 + 9\alpha + n - 3 > \Pi_R^I(\tilde{n}_R,3,3) = -\frac{9}{2}\alpha^2 + (\tilde{n}_R+6)\alpha + n - \tilde{n}_R$, which shows that the payoff at the SPNE with $n_R = 3$ is greater than that at the SPNE with $\tilde{n}_R$ for all countries in $\tilde{P}$. We have $\Pi_F^O(3,3,3) = 9\alpha + n - \frac{7}{2} > \Pi_F^O(\tilde{n}_R,3,3) = (\tilde{n}_R+6)\alpha + n - \tilde{n}_R - \frac{1}{2}$, which shows that the payoff at the SPNE with $n_R = 3$ is greater than that at the SPNE with $\tilde{n}_R$ for all countries that choose F in both equilibria. Finally, we have $\Pi_F^O(3,3,3) = 9\alpha + n - \frac{7}{2} > \Pi_R^O(\tilde{n}_R,3,3) = -\frac{1}{2}\alpha^2 + (\tilde{n}_R+6)\alpha + n - \tilde{n}_R$, which shows that the payoff at the SPNE with $n_R = 3$ is greater than that at the SPNE with $\tilde{n}_R$ for all countries that choose F in the SPNE with $n_R = 3$ and R in that with $\tilde{n}_R$. ||

In conclusion, $n_R = 3$ is the one and only number attained at PU-SPNE in Case 2. ■

Proof of Result 3

We show that $\Pi_F^I(3,n,3) < \Pi_R^I(4,n,4)$. By (12) and (14):

$$\Pi_F^I(3,n,3) = \frac{-15+9n}{2n}\alpha^2 + \frac{36-24n+6n^2}{2n}\alpha + \frac{-21+15n-6n^2+n^3}{2n} \text{ and}$$
$$\Pi_R^I(4,n,4) = \frac{-28+23n-8n^2}{2n}\alpha^2 + \frac{64-56n+26n^2-2n^3}{2n}\alpha + \frac{-36+33n-18n^2+3n^3}{2n}.$$

Thus, we have:

$$\Pi_R^I(4,n,4) - \Pi_F^I(3,n,3) = -\frac{13-14n+8n^2}{2n}\alpha^2 - \frac{28+32n-20n^2+2n^3}{2n}\alpha - \frac{15-18n+12n^2-2n^3}{2n}.$$

We find that $\Pi_R^I(4,n,4) - \Pi_F^I(3,n,3) > 0$ for all α such that $0 < \alpha < 1$ because (i) the coefficient of α^2 is negative if $n \geq 5$, (ii) $\Pi_R^I(4,n,4) - \Pi_F^I(3,n,3) > 0$ if $\alpha = 0$ and $n \geq 5$, and (iii) $\Pi_R^I(4,n,4) - \Pi_F^I(3,n,3) = 0$ if $\alpha = 1$. ■

F Partial delegation

We consider two additional types of delegation. One is the *participation-decision delegation* (*PD-delegation*), in which the FG delegates only the decision to participate in the negotiation. In this type of delegation, all decisions in Stage 3 are made by the FG. Moreover, if the RG of a country decides that the country should participate in a negotiation, then the FG negotiates and receives international transfers. The other type of delegation we consider here is the *negotiation delegation* (*negot-delegation*), in which the FG only delegates decisions in Stage 3. Thus, the participation decision in Stage 2 is made by the FG, but if the FG decides in favor of participating in a negotiation, then the RG negotiates and receives transfers from other countries. These two delegations are considered to be intermediate between actions F (*no delegation*) and R (*full delegation*) in the main text.

The new types of delegation capture some real-world situations. In some international negotiations, the executive of a country, whose preferences are based on the country's welfare, negotiates with other countries and makes international treaties. The treaties are effective subject to ratification by legislators, whose preferences are based on the welfare of their local constituencies. This would be captured by the PD-delegation. To understand the negotiation delegation, imagine a different situation, in which a country recognizes the need for international cooperation to combat some environmental problem. Despite this national recognition, the country faces significant internal conflicts of interest between the subnational regions that are difficult to reconcile. In this case, it would be an option to delegate only the negotiation to the regional representatives.[7)]

F.1 Payoffs under participation-decision delegation (PD-delegation)

As in the main text, we examine the case of $v(g) = g$ and $c(g) = \frac{g^2}{2}$. If country $i \in N$ chooses the PD-delegation in Stage 1, then the RG of the polluter region makes the participation decision and the FG is the negotiator when the RG chooses participation. We assume that the FG makes the decision in Stage 3 even if the negotiation breaks down or country i does not participate in the negotiation. Because the FG is the negotiator, it receives the transfers:

$$T_i^P = -\left(v\Big(\sum_{j\in P} g^P(p, p_R)\Big) - c(g^P(p, p_R))\right) + v\Big(\sum_{j\in P} g_j^d\Big) - c(g_F^d) + \frac{S(p, p_R)}{p},$$

when P is a set of participating countries with (p, p_R) such that $i \in P$. Given the participation of $P \setminus \{i\}$ with $(p-1, p_R)$, if the RG chooses participation in the negotiation, it obtains the payoff:

$$\alpha v\Big(\sum_{j\in P} g^P(p, p_R) + \sum_{j\in N\setminus P} g_j^d\Big) - c(g^P(p, p_R)). \tag{17}$$

Conversely, if it chooses nonparticipation, then it obtains the payoff:

$$\alpha v\Big(\sum_{j\in P\setminus\{i\}} g^P(p-1, p_R) + \sum_{j\in N\setminus P} g_j^d + g_F^d\Big) - c(g_F^d). \tag{18}$$

The RG chooses participation if (17) $\geq$ (18). If the FG chooses the PD-delegation, then its payoff is $v\left(\sum_{j\in N} g_j^d\right) - c(g_F^d) + \frac{S(p,p_R)}{p}$ if the RG chooses participation and $v\Big(g_F^d + \sum_{j\in P\setminus\{i\}} g^P(p-1, p_R) + \sum_{j\in N\setminus P} g_j^d\Big) - c(g_F^d)$ otherwise.

[7)] This situation resembles the establishment of the Pacific Salmon Treaty, in which some US state governments participated in the international negotiations. See Yanagida (1987) for details.

F.2 Payoffs under delegation of the negotiation (Negot-delegation)

If country $i \in N$ chooses negot-delegation in Stage 1, then the FG makes the participation decision and the RG is the negotiator when the FG chooses participation. We assume that the RG makes the decision in Stage 3 even if the negotiation breaks down or country i does not participate in the negotiation. Because the RG is the negotiator, it receives the transfers:

$$T_i^P = -\left(\alpha v\Big(\sum_{j\in P} g^P(p, p_R)\Big) - c(g^P(p, p_R))\right) + \alpha v\Big(\sum_{j\in P} g_j^d\Big) - c(g_R^d) + \frac{S(p, p_R)}{p},$$

when P is a set of participating countries with (p, p_R) such that $i \in P$. Given the participation of $P \setminus \{i\}$ with $(p-1, p_R-1)$, the FG obtains the following payoff if it chooses participation:

$$\begin{aligned} &v\Big(\sum_{j\in P} g^P(p, p_R) + \sum_{j\notin P} g_j^d\Big) - c(g^P(p, p_R)) + T_i^P \\ &= (1-\alpha)\Big(\sum_{j\in P} g^P(p, p_R) + \sum_{j\in N\setminus P} g_j^d\Big) + \alpha\Big(\sum_{j\in P} g_j^d + \sum_{j\in N\setminus P} g_j^d\Big) - c(g_R^d) + \frac{S(p, p_R)}{p}. \end{aligned} \tag{19}$$

If it chooses nonparticipation, the payoff is:

$$\sum_{j\in P\setminus\{i\}} g^P(p-1, p_R-1) + \sum_{j\notin P} g_j^d + g_R^d - c(g_R^d). \tag{20}$$

The FG chooses participation if (19) $\geq$ (20).

F.3 Results

The introduction of the two new types of delegation does not change Proposition 1 in the main text. Result 4 shows that every FG makes the participation decision and negotiates by itself in a SPNE.

Result 4 *There always exists a SPNE at which all FGs choose no delegation in Stage 1.*

Proof. We show that the following strategy is subgame perfect. Let $P^* \equiv \{1, 2, 3\} \subseteq N$:

Stage 1. All FGs choose no delegation.

Stage 2. Depending on the choice in Stage 1, the participation behavior of each country is defined for all subgames starting from Stage 2 as follows.

(2.1) If all FGs choose no delegation in Stage 1, then P^* is the set of participants.

(2.2) If the FG of one of the countries (e.g., $i \in N$) chooses PD-delegation and the others choose no delegation in Stage 1, then $P^{(2.2)} \equiv \{j, k\}$ such that $j, k \neq i$ is the set of participants.

(2.3) If the FG of one of the countries (e.g., $i \in N$) chooses negot-delegation and the others choose no delegation in Stage 1, then $P^{(2.3)} \equiv \{i, j\}$ such that $j \neq i$ is the set of participants.

(2.4) If the FG of one of the countries (e.g., $i \in N$) chooses full delegation and the others choose no delegation in Stage 1, then $P^{(2.4)} \equiv \{i, j\}$ such that $j \neq i$ is the set of participants.

(2.5) Otherwise, the set of participants is any set supported at a Nash equilibrium.

In the strategy, every FG of the country in P^* obtains the payoff $v(ng_F^d) - c(g_F^d) + \frac{S(3,0)}{3} = n + \frac{3}{2}$ and the FG of the country outside P^* obtains the payoff $v(3g^P(3,0) + (n-3)g_F^d) - c(g_F^d) = n + \frac{11}{2}$. Note that $n + \frac{11}{2} > n + \frac{3}{2}$.

First, we have already shown that P^* is a Nash equilibrium set of participants in Stage 2 after all FGs choose no delegation (see Claim 1 in the main text). We have also shown that if country i's FG chooses full delegation and the others choose no delegation, then $P^{(2.4)}$ ($j \neq i$) is a Nash equilibrium set of participants in Stage 2 (see Claim 2 in the main text).

Claim F.1 *In Stage 2 after country i's FG chooses PD-delegation and the others choose no delegation, $P^{(2.2)} = \{j, k\}$ such that $j, k \neq i$ is a Nash equilibrium set of participants.*

Proof of Claim F.1 Note that the participation decision of every country in $P^{(2.2)}$ is made by the FG. No countries in $P^{(2.2)}$ are made better off by deviating to nonparticipation by the property of the NBS (see $ISF(2,0) > 0$ in Lemma B.2-(a)).

If the RG of country i chooses to participate in the negotiation, then its payoff is $\alpha v(3g^P(3,0) + (n-3)g_F^d) - c(g^P(3,0)) = (n+6)\alpha - \frac{9}{2}$ by (17). If it chooses not to participate, then its payoff is $\alpha v(2g^P(2,0) + (n-2)g_F^d) - c(g_F^d) = (n+2)\alpha - \frac{1}{2}$ by (18). Subtracting the latter from the former yields $4(\alpha - 1) < 0$.

The participation decisions of countries other than i, j, and k are made by the FGs. In the strategy, they choose nonparticipation. As $ISF(3,0) = 0$, deviating to participation would not make them better off (see Lemma B.2-(b)). Thus, $P^{(2.2)}$ is a Nash equilibrium set of participants. ||

Claim F.2 *In Stage 2, after country i's FG chooses negot-delegation and the others choose no delegation, $P^{(2.3)} = \{i, j\}$ such that $j \neq i$ is a Nash equilibrium set of participants.*

Proof of Claim F.2 By Lemma B.2-(a), $ISF(2,1) > 0$, which implies that country j's FG does not deviate from participation to nonparticipation.

By (19) and (20), country i's FG obtains the payoff $(1-\alpha)[2g^P(2,1)+(n-2)g_F^d]+\alpha[g_R^d+(n-1)g_F^d]-c(g_R^d)+\frac{S(2,1)}{2} = -\frac{5}{4}\alpha^2+\frac{3}{2}\alpha+n+\frac{1}{4}$ if it chooses participation and the payoff $g_R^d+(n-1)g_F^d-c(g_R^d) = \alpha+n-1-\frac{\alpha^2}{2}$ if it chooses nonparticipation. Subtracting the latter payoff from the former yields $\frac{1}{4}(-3\alpha^2+2\alpha+5)$, which is positive because $0 < \alpha < 1$. Thus, country i's FG does not deviate from participation to nonparticipation.

Finally, by Lemma B.2-(b), $ISF(3,1) = \frac{1}{3}(\alpha-1)\alpha < 0$, which implies that country $k \neq i, j$ does not deviate from nonparticipation to participation. Hence, $P = \{i, j\}$ is a Nash equilibrium set of participants. ||

Finally, we examine Stage 1. If country i unilaterally deviates to PD-delegation, then country i's FG obtains the payoff $v(2g^P(2,0) + (n-2)g_F^d) - c(g_F^d) = n + \frac{3}{2}$. If country i unilaterally deviates to negot-delegation, then country i's FG obtains the payoff:

$$(1-\alpha)\left(2g^{P^{(2.3)}}(2,1) + g_F^d\right) + \alpha\left(2g_F^d + g_R^d\right) - c(g_R^d) + \frac{S(p,1)}{2} = -\frac{5}{4}\alpha^2 + \frac{3}{2}\alpha + n + \frac{1}{4}.$$

As in (3) in the main text, if country i's FG unilaterally deviates to full delegation, then it obtains the payoff $-\frac{5}{4}\alpha^2 + \alpha + n + \frac{1}{4}$. When country i's FG deviates, none of the payoffs are greater than $n + \frac{3}{2}$ for cases $i \in P^*$ and $i \notin P^*$. Thus, all FGs choose nondelegation in a SPNE. ■

As in the proof of Proposition 1, the equilibrium set of participants shrinks if one FG deviates in Stage 1 in the proof of Result 4. The two delegation options of full delegation and negot-delegation

generate a strategic advantage in a negotiation if the countries choosing these options participate in the negotiation. Given this participation, these options have the same impact on the other countries choosing no delegation because the negotiators under these options are the RGs. Thus, the discussion after Proposition 1 in the main text applies: that is, the country choosing no delegation does not participate in the negotiation when manipulation through delegation is effective (i.e., α is high enough). Because of this, the equilibrium set of participants shrinks, and the strategic advantage achieved through full delegation and negot-delegation is reduced.

References in the online appendix

[1] D'Aspremont, C.A., Jacquemin, J., Gabszeweiz, J., Weymark, J.A., 1983. On the stability of collusive price leadership. *Canadian Journal of Economics* 16, 17–25.

[2] Eckert, H., 2003. Negotiating environmental agreements: Regional or federal authority? *Journal of Environmental Economics and Management* 46, 1–24.

[3] Yanagida, J.A., 1987. The Pacific Salmon Treaty. *American Journal of International Law* 81 (3), 577–592.