When does Disinformation Promote Successful Treaties?

Ana Espínola-Arredondo*
School of Economic Sciences
Washington State University
Pullman, WA 99164

Félix Muñoz-García†
School of Economic Sciences
Washington State University
Pullman, WA 99164

March 7, 2012

Abstract

This paper analyzes the negotiation of a non-binding treaty where agents are uninformed about each others' ability to comply with the terms of the agreement. We show that the presence of incomplete information allows the treaty to become successful under larger set of parameter conditions than under complete information. The paper also examines the welfare properties of our equilibrium results, showing a welfare improvement relative to complete information under certain conditions. Finally, we extend our analysis to settings where countries' types are correlated, finding that the equilibrium outcome where information is conveyed can be supported under larger conditions.

KEYWORDS: Signaling games; Information transmission; Non-binding negotiations; Correlated types.

JEL CLASSIFICATION: C72, D62, Q28.

*Address: 111C Hulbert Hall, Washington State University, Pullman, WA 99164. E-mail: anaespinola@wsu.edu.

†Address: 103G Hulbert Hall, Washington State University. Pullman, WA 99164-6210. E-mail: fmunoz@wsu.edu. Phone: (509) 335 8402. Fax: (509) 335 1173.

1 Introduction

The United States leading role in the 2009 United Nations Climate Change Conference raised questions about the U.S. true commitment to the content of the agreement. On one hand, the then-recent election of President Obama led many countries to speculate that the U.S. environmental agenda could experience a change of course, relative to the previous administration, given the strong emphasis on environmental issues by the Obama's presidential campaign.[1] On the other hand, several specialists questioned the U.S. commitment on the grounds that it did not ratify the Kyoto's protocol and the current difficulty of passing a climate change bill through a reluctant Senate.[2] A similar concern was raised about China's true commitment to environmental policies. Specifically, China's President Hu Jintao promised a 40-45 percent cut in emissions by 2020, but did not agree to outside verification, leading many observers to express their skepticism about China's fulfillment of its promises.[3]

Besides environmental treaties, noncompliance has also been observed in other international agreements. For instance, in the context of trade agreements, Busch and Reinhardt (2003) found that from 1948 to 2002 the U.S. faced 193 disputes in the GATT/WTO, where 56 rulings were against this country, indicating that the U.S. did not comply with some requirements of the treaty. Noncompliance has additionally been documented in security agreements, Martin (2005). For example, the U.S. has maintained a large-scale development of Anti-Ballistic Missiles (ABMs) despite of its 1972 ABMs Treaty with the Soviet Union, where both countries agreed to limit the construction of ABM installations. Similarly, the U.S. signed the Comprehensive Nuclear-Test-Ban Treaty (CTBT) of 1996 during the Clinton administration, but afterwards it was not ratified by a Republican dominated Senate in 1999.[4,5]

Countries, hence, frequently face uncertainty about whether or not the content of the agreement will be respected by other signatories. In order to model this information setting, our paper considers a bilateral negotiation where the representatives of both countries are privately informed about their own type, where countries' types capture any underlying characteristic affecting their political ability to comply with the terms of the treaty. Specifically, when the type of a country's representative is "high," he complies with the content of the agreement since he has the political

[1] For instance, PolitiFact.com tracked over 53 different promises related to environmental and energy issues during the Obama campaign.

[2] During the Copenhagen summit Sweden's minister of environment, Andreas Carlgren, was one of the leaders expressing his worries about the U.S. willingness to commit to policies curving global warming; as reported in the New York Times, December 26, 2009. In this line, according to Schreurs et. al. (2009), pp. 8-9, the U.S. signed nine environmental treaties over the past 30 years, but they did not receive enough votes for Senate ratification.

[3] As reported in the New York Times last November 26, 2009.

[4] Most senators voted along party lines, with 92 percent of Republican senators voting against the treaty, and 100 percent of Democrats voting in favor. For more details, see U.S. Senate Rollcall Vote No. 325 (October 13, 1999).

[5] Examples abound of treaties signed by a country but subsequently not ratified by its Senate. For instance, the U.S. played a leading role and signed the League of Nations of 1919 but, due to opposition in its Senate, never ratified the Covenant. Likewise, the U.S. signed the United Nations Convention on the Law of the Sea (UNCLOS) of 1994 but did not ratify it either. Finally, 20 E.U. states signed the European Convention on Human Rights and Biomedicine in 1997-98, but it has not yet been ratified by Finland, France, Italy, Latvia, Luxembourg, The Netherlands, Poland and Sweden; as documented by the Council of Europe (see Concil of Europe Treaty Series No. 164).

power, both within his party and the Congress, to implement the treaty. In contrast, when the type of the country's representative is "low," he lacks such political power and, therefore, does not fully comply with the terms of the agreement. A representative's political ability to fully comply with the content of the agreement is unobservable by representatives from other countries. Indeed, a country's official can approximately estimate other officials' domestic political power. Yet, he cannot perfectly assess the relationships these officials maintain across party lines, or other informal associations between the main political parties, which ultimately affect the posterior implementation of the agreement. Finally, since our model allows for both countries to be uninformed about each others' types, the paper also embodies standard signaling games with one-sided uncertainty as special cases, as well as negotiation games with perfectly informed players.

The paper analyzes a signaling game where the country leading the negotiation decides whether to participate in the agreement and, conditional on the leader's signature, the following country chooses whether to join the treaty. We first identify a fully informative separating equilibrium, where only the high-type leader signs the agreement, thereby conveying its political ability to the uninformed follower, which responds joining the treaty regardless of its type. Moreover, a pooling equilibrium emerges where information about the leader's type is concealed from the follower. This equilibrium can be sustained if both leader's and follower's priors are sufficiently high and, therefore, both countries sign regardless of their types. For instance, Japan's participation in international whaling agreements might be explained by this equilibrium prediction. Despite its previous history, Japan led the negotiations of a bilateral agreement with the U.S. by which Japan committed in 1984 to end all whaling by 1988. However, Japan continued its whaling practices afterwards, claiming such whaling was "research" oriented, and hence allowed under the International Whaling Commission's rules.[6]

In addition to these two pure-strategy equilibria, we also describe under which conditions a semi-separating equilibrium arises where both countries randomize their participation decision. Hence, relative to a complete information benchmark, whereby the agreement is only successful when the leader's type is high, our equilibrium results suggest that treaties become successful under conditions for which the agreement would not be signed in complete information contexts. Thus, the presence of incomplete information, rather than hindering the chances of a successful agreement, can actually promote its signature.[7]

We then examine the welfare properties of our equilibrium results. In particular, we demonstrate that, when the type of at least one of the countries involved in the negotiations is high, social welfare in the pooling equilibrium —where both countries sign— is weakly larger than under any of the other equilibrium outcomes, where either: only one type of leader participates (as in the separating equilibrium), or the leader randomizes its participation decision (as in the semiseparating equilibrium). In contrast, when both countries' types are low, the no signature of the agreement prescribed in the separating equilibrium yields a larger social welfare. Therefore, from a policy

[6] For more details on this whaling agreement, among others, see Stoett (1997).

[7] This result is in the line with Young (1991), as opposed to Keohane (1984) who argues that more information facilitates international cooperation.

perspective, our results suggest that international organizations should favor negotiations where at least one country has a relatively consistent history of compliance with the content of similar treaties. We also compare social welfare under complete and incomplete information settings. Specifically, we show that the introduction of incomplete information leads to a welfare improvement if the type of at least one country is high and, in addition, countries' priors are relatively symmetric. Otherwise, uncertainty does not necessarily yield a larger social welfare.

Finally, we investigate how our equilibrium outcomes are affected by the presence of correlation in countries' types. This context might arise when countries share similar institutional settings and political scenarios, thereby increasing the likelihood that their types coincide. We demonstrate that positive correlation expands (shrinks) the region of parameter values under which the separating (pooling, respectively) equilibrium can be supported. Importantly, since the pooling equilibrium entails a larger social welfare than under complete information, our results suggest that, from a policy perspective, international conferences should promote settings where countries' types actually differ, since they increase the likelihood of a welfare-improving agreement.

Our model can be extended to settings where domestic agents —such as unions, political parties, etc.— negotiate whether to accept an agreement proposed by an independent party, e.g., the Congressional Budget Office, and are uncertain about each others' ability to fulfill the content of the treaty. This paper suggests that the presence of uncertainty into these types of negotiations can actually entail an increase in the probability that the agreement is successful which, in some cases, can provide welfare benefits.

Related literature. The literature on international agreements has extensively examined negotiations under a context of complete information; see Barrett (1994a and 1999) and Cesar (1994). These studies show that when free-riding incentives are small treaties can be sustained under larger parameter conditions.[8] Several international agreements are, however, usually negotiated in contexts of incomplete information, like that analyzed in our model. This paper hence contributes to the literature considering uncertainty in international negotiations, such as Iida (1993), who analyzes treaties using a repeated bargaining game. Specifically, he assumes that a country is uninformed about other countries' domestic constraints, whereas we consider that both countries are uninformed about each others' ability to comply with the content of the treaty.[9] Martin (2005) also analyzes the signaling role of the signature of a treaty. Her paper considers two different types of agreements, executive treaties and international agreements, which imply different degrees of compliance. Unlike her study, we investigate the case where not only the follower but also the

[8] This literature was extended by models allowing countries to impose "sanctions" on defecting countries, Barrett (1994b), and by studies linking the negotiations of transboundary pollution treaties with other issues such as free-trade agreements; see Whalley (1991), Carraro and Siniscalco (2001) and Ederington (2002). Fearon (1998) studies a repeated game where countries, despite being perfectly informed, cannot perfectly detect each others' defection from the cooperative agreement.

[9] Tarar (2001) extends Iida's (1993) model to a setting where the executive of one country in the negotiation observes both his own domestic constraint and that of the other country. (For similar models on unilateral uncertainty, see Espinola-Arredondo and Munoz-Garcia (2011).) Our model, however, considers that both executives are privately informed about their domestic constraint, and allows for correlation.

leader is uninformed. Our paper also differs along several other dimensions: we allow for a more general payoff structure, provide an analysis of settings where players' types are correlated, and compare the welfare properties of different information contexts.[10]

Koremenos (2005) empirically analyzes states' willingness to participate in binding agreements when countries can suffer future shocks affecting the benefits of the treaty. She shows that countries, in order to insure themselves from uncertainty, sign finite duration agreements and include provisions that allow for future renegotiations.[11] Koremenos' (2005) study contributes to a recent line of work, often regarded as the "rational design" literature, which examines how institutional design is affected by the conditions surrounding the negotiations, such as the availability of information and the possibility of monitoring.[12] In particular, these studies assume that countries are uncertain about their own future political or economic scenario, and thus hesitate to participate in binding agreements. Negotiating countries are therefore exposed to an ex-ante symmetric uncertainty, in the form of a probabilistic future shock, when deciding whether to sign the treaty. In contrast, we consider that the uncertainty country representatives face is asymmetric, namely, a country's executive can accurately assess his own political ability to implement the content of the treaty, yet cannot correctly estimate the ability of his cosignatories.

The next section describes the model under complete information. Section 3 examines the set of equilibria when countries are privately informed, their comparative statics and the cases that arise when only one of the countries is privately informed about its type. Section 4 provides welfare comparisons. In section 5 we extend our analysis to correlated types and offer policy implications, and section 6 concludes.

2 Model

Let us examine an international treaty as a game where countries' representatives decide whether or not to participate in the agreement. The country acting as the leader (country 1) announces whether it joins the agreement. If the leader signs the treaty, then the follower (country 2) decides whether or not to sign. The commitment levels specified in the treaty are non-binding and its implementation provide benefits to all countries, such as a limit in the production of anti-ballistic missiles, or the reduction of pollution in an environmental treaty. However, the leader obtains a larger benefit from the treaty than the follower, and hence the former has incentives to carry the burden of the negotiations. For instance, it is more affected by environmental pollution, or it is more threatened by other countries with nuclear capabilities.

In this section, we consider a complete information setting where countries observe each others'

[10] In a different context, the literature on international trade has recently examined tariff agreements where countries are privately informed, for instance, about the extent to which the import-competing sector of another country is affected by an efficiency shock; see Lee (2007), Martin and Vergote (2008) and Bagwell (2009).

[11] Likewise, Von Stein (2008) empirically studies how the introduction of flexibility provisions in international environmental agreements affects their ratification.

[12] See, among others, Downs et al. (1996), Koremenos (2001), Koremenos et al. (2004), Lake (1999), Rosendorff (2005) and Yarbrough and Yarbrough (1992).

types: high, θ_H, implying that the country has the political ability to comply with the terms of the treaty; or low, θ_L, which indicates that it behaves as if the treaty was not signed (status quo, where countries select Nash equilibrium strategies). If the treaty is signed by both countries, it becomes successful, and a signatory complies with the content of the agreement if its type is high, whereas it behaves as a non-signatory if its type is low. If either country does not sign the treaty, it becomes unsuccessful and both countries behave as in the status quo. In particular, let country i's equilibrium payoff from the leader's signature be represented by $V_{i_K}(S, s_j; \theta_J)$, where the first term of the subscript denotes the country's position in the negotiation, $i = \{1, 2\}$; the second term represents its type, $K = \{H, L\}$; and θ_J denotes the type of country $j \neq i$, where $J = \{H, L\}$. In addition, $s_j = \{S, NS\}$ represents the follower's strategy (the signature or not signature of the treaty, respectively), when the leader signs the agreement. For compactness, let $V_{i_K}(NS; \theta_J)$ denote country i's payoff in the case that the leader does not sign the agreement, and the follower's action set is therefore empty.

Leader's payoffs. When the first mover does not participate, the treaty is not successful and its equilibrium payoff is $V_{1_K}(NS; \theta_J)$. When the leader signs the treaty but the follower responds not participating, the leader's equilibrium payoff becomes $V_{1_K}(S, NS; \theta_J)$. In this setting, the agreement is still not successful either, and countries' actions coincide with those under no treaty (status quo). Nonetheless, the leader bears a cost from a failed agreement. Hence, if the follower does not participate, the leader prefers not to sign the treaty, i.e., $V_{1_K}(NS; \theta_J) > V_{1_K}(S, NS; \theta_J)$, since by not signing the leader avoids any cost from an unsuccessful treaty. When the follower responds joining, the leader's equilibrium payoff is $V_{1_K}(S, S; \theta_J)$. Therefore, $BS_{1_K}(\theta_J) \equiv V_{1_K}(S, S; \theta_J) - V_{1_K}(NS; \theta_J)$ describes the leader's benefit from signing a treaty with a θ_J-type follower. When the leader's type is high, the benefits from the signature (e.g., improved environmental quality) offset its associated cost, regardless of the follower's type, i.e., $BS_{1_H}(\theta_H) > BS_{1_H}(\theta_L) > 0$. In contrast, when the leader's type is low, it prefers to avoid an agreement with a low-type follower, i.e., $BS_{1_L}(\theta_H) > 0 > BS_{1_L}(\theta_L)$.[13] The following example illustrates these conditions on the leader's payoffs in the context of public good games.

Example. Similar to standard public good games, consider that the K-type country payoff function is given by

$$U_i^K(e_i, e_j) = Benefit_i^K(e_i, e_j) - Cost_i(e_i; \theta_K) = \ln\left[\alpha_i m(e_i + e_j)\right] - \frac{e_i}{\theta_K}, \quad \text{where } j \neq i$$

and e_i denotes the effort that country $i = \{1, 2\}$ exerts implementing the content of the treaty, e.g., its investment in clean technologies in an environmental agreement, and $\alpha_1 = 2$ for the leader and $\alpha_2 = 1$ for the follower. As suggested above, if the K-type leader does not sign the treaty, the negotiation game ends, and every country i independently and simultaneously solves $\max_{e_i}$

[13] This indicates that the negotiation cost that the low-type leader must incur outweighs the benefits from signing a treaty with a low-type follower, since no signatory complies with the content of the agreement, ultimately yielding a negative benefit from the agreement.

$U_i^K(e_i, e_j)$, yielding the Nash equilibrium effort levels $\left(e_i^{NE}, e_j^{NE}\right) = \left(\frac{\theta_K}{2}, \frac{\theta_K}{2}\right)$ when both countries' types coincide, and $\left(e_i^{NE}, e_j^{NE}\right) = (\theta_H, 0)$ when countries' types do not coincide and the type of country $i \neq j$ is high. In this context, the leader's equilibrium payoff becomes $V_{1_K}(NS; \theta_K) = \ln[2m\theta_K] - \frac{1}{2}$ when both countries' type is $K = \{H, L\}$, it is $V_{1_H}(NS; \theta_L) = \ln[2m\theta_H] - 1$ when the leader's (follower's) type is high (low, respectively), and $V_{1_L}(NS; \theta_H) = \ln[2m\theta_H]$ if only the follower's type is high. When the leader signs the treaty but the follower responds not joining, the treaty is also unsuccessful and each country i solves $\max_{e_i} U_i^K(e_i, e_j)$ by selecting the Nash equilibrium, entailing an equilibrium payoff of $V_{1_K}(S, NS; \theta_J) = V_{1_K}(NS; \theta_J) - NC$, which lies below $V_{1_K}(NS; \theta_J)$, given that the leader must incur the negotiation cost $NC \in \mathbb{R}_+$. If, instead, the follower responds joining, the treaty is successful. In this setting, the social optimum solves

$$\max_{e_1, e_2} U_1^K(e_1, e_2) + U_2^J(e_1, e_2).$$

When both countries' types are high, $\left(e_i^{SO}, e_j^{SO}\right) = (\theta_H, \theta_H)$, while when they are low, they independently choose effort level $\frac{\theta_L}{2}$ as under no treaty. However, when their types differ, $\left(e_i^{SO}, e_j^{SO}\right) = (2\theta_H, 0)$ when only country i's type is high. In this setting, the leader's equilibrium payoff from signing the treaty is thus $V_{1_H}(S, S; \theta_H) = \ln[4m\theta_H] - 1 - NC$ when both countries' types are high, $V_{1_L}(S, S; \theta_L) = \ln[2m\theta_L] - \frac{1}{2} - NC$ when both types are low, $V_{1_H}(S, S; \theta_L) = \ln[4m\theta_H] - 2 - NC$ if only the leader's type is high, and $V_{1_L}(S, S; \theta_H) = \ln[4m\theta_H] - NC$ if only the follower's type is high. Therefore, the high-type leader's benefit from signing the treaty is $BS_{1_H}(\theta_H) = 0.89 - NC$ and $BS_{1_H}(\theta_L) = 0.39 - NC$, i.e., $BS_{1_H}(\theta_H) > BS_{1_H}(\theta_L) > 0$ if and only if $NC < 0.39$. By contrast, the low-type leader's benefit is $BS_{1_L}(\theta_H) = 0.69 - NC$ and $BS_{1_L}(\theta_L) = -NC$, i.e., $BS_{1_L}(\theta_H) > 0 > BS_{1_L}(\theta_L)$.

Follower's payoffs. The K-type follower obtains a payoff $V_{2_K}(NS; \theta_J)$ when the θ_J-type leader does not participate. If, instead, the leader signs the agreement but the follower responds not participating, its payoff is $V_{2_K}(S, NS; \theta_J)$. When the follower responds signing the treaty its payoff becomes $V_{2_K}(S, S; \theta_J)$. Similarly as for the leader's payoff, let $BS_{2_K}(\theta_J) \equiv V_{2_K}(S, S; \theta_J) - V_{2_K}(S, NS; \theta_J)$ denote the K-type follower's benefit from joining a treaty with a θ_J-type leader. Assume that the high-type follower's incentives to participate in the agreement are positive when facing a high-type leader, i.e., $BS_{2_H}(\theta_H) > 0$, but negative otherwise, $BS_{2_H}(\theta_L) < 0$. Intuitively, the high-type follower benefits more from a larger compliance of the treaty when the leader is also a high-type country than when the leader is not, whereby the fulfillment of the agreement is mainly borne by the follower. Furthermore, $BS_{2_H}(\theta_L)$ is negative since the low-type leader does not comply and, hence, the improvement in the global environmental quality does not compensate the increase in abatement costs that the follower experiences. When the follower is a low type, it benefits from participating in a treaty with the high-type leader (where it free-rides the leader's compliance), i.e., $BS_{2_L}(\theta_H) > 0$, but does not obtain any benefit from signing an agreement with

the low-type leader, i.e., $BS_{2_L}(\theta_L) = 0$.[14]

Example. Following the above example, when the leader does not sign the treaty, the follower's equilibrium payoff is $V_{2_K}(NS;\theta_K) = \ln[m\theta_K] - \frac{1}{2}$ if both countries' types coincide, and $V_{2_H}(NS;\theta_L) = \ln[m\theta_H] - 1$ when the follower's (leader's) type is high (low, respectively), and $V_{2_L}(NS;\theta_H) = \ln[m\theta_H]$ if only the leader's type is high. When the leader signs but the follower responds not joining, the follower's equilibrium payoff is $V_{2_K}(S, NS;\theta_J)$ which coincides with $V_{2_K}(NS;\theta_J)$. By contrast, if the follower responds joining, the treaty becomes successful, implying that the follower's equilibrium payoff from signing the treaty is thus $V_{2_H}(S, S;\theta_H) = \ln[2m\theta_H] - 1$ when both countries' types are high, $V_{2_L}(S, S;\theta_L) = \ln[m\theta_L] - \frac{1}{2}$ when both types are low, $V_{2_H}(S, S;\theta_L) = \ln[2m\theta_H] - 2$ if only the follower's type is high, and $V_{2_L}(S, S;\theta_H) = \ln[2m\theta_H]$ if only the leader's type is high. Hence, the high-type follower's benefit from joining the treaty is $BS_{2_H}(\theta_H) = 0.19$ and $BS_{2_H}(\theta_L) = -0.31$, i.e., $BS_{2_H}(\theta_H) > 0 > BS_{2_H}(\theta_L)$. However, the low-type follower's benefit from the agreement is $BS_{2_L}(\theta_H) = 0.69 > 0$ and $BS_{2_L}(\theta_L) = 0$.

We next identify the equilibrium strategy in the subgame perfect equilibrium of the game.

Lemma 1 (Complete information). *The leader signs (does not sign) the treaty when its type is high (low, respectively), regardless of the follower's type. The follower responds joining the treaty when he observes that a high-type leader participates, but does not sign the agreement otherwise.*

Intuitively, the leader anticipates that the follower will only respond joining the treaty when it observes a signature from a high-type leader. As a consequence, the high-type leader signs the agreement, whereas the low-type leader does not participate since the follower would respond not joining it. Therefore, under complete information, the treaty is only successful when the leader's type is high. In the following section, we investigate if the introduction of incomplete information about countries' types allows for the emergence of equilibrium outcomes where countries sign under conditions for which the treaty would not be successful under complete information, namely, when the leader's type is low.

3 Signaling game

In this section we consider settings where every country is privately informed about its own type, but does not observe the other country's type. This case describes strategic contexts where both leader's and follower's commitment with the agreement is uncertain because, for instance, both countries have a previous history of irregularly complying with the terms of the treaty, or both countries have recently elected officials. Specifically, the time structure of the signaling game is as follows:

[14] Note that a negotiation between a low-type leader and follower implies that countries select the same second-period actions whether or not the treaty is signed (they behave as in the status quo). Hence, the benefits from participating in the treaty coincide with those of not participating.

1. Nature selects the leader's type, which is privately observed by the leader but not by the follower. For simplicity, assume that the leader's type is either high, θ_H, or low, θ_L, with associated probabilities p and $1-p$, respectively. Similarly, nature determines the follower's type, which is also privately observed by the follower, either high or low, with probabilities q and $1-q$, respectively.[15]

2. After observing its type, the leading country announces its participation in the agreement. If the leader does not sign the treaty, the negotiation ends, and the agreement is unsuccessful, and every country behaves as in the status quo.

3. After observing the leader's signature, the J-type follower forms posterior beliefs about the leader's type being high, i.e., $\mu_J(H|S)$ where $J = \{H, L\}$.[16] Given its posterior beliefs, the follower chooses to sign or not sign the agreement. If the follower responds not participating, the negotiation ends.

4. If the treaty is signed by both players, high-type countries fully implement the content of the agreement, while those with a low type implement the status quo.[17]

We next study under which conditions the leader's decision to sign the agreement conveys or conceals information about its type to the follower, thus affecting the follower's willingness to participate in the treaty. The following proposition describes the set of Perfect Bayesian equilibria (PBE) in pure strategies in this signaling game.

Proposition 1. *In the signaling game where all countries are uninformed about each others' types, the following equilibria in pure strategies survive the Cho and Kreps' (1987) Intuitive Criterion:*

a. *A separating equilibrium in which the leader signs (does not sign) the treaty when its type is high (low, respectively), and the follower responds participating (not participating) after observing a signature (no signature, respectively), both when its own type is high and low, if and only if* $q < \overline{q}_L$, *where* $\overline{q}_L \equiv \frac{-BS_{1_L}(\theta_L)}{BS_{1_L}(\theta_H)-BS_{1_L}(\theta_L)}$; *and*

b. *A pooling equilibrium in which both types of leader sign and both types of follower respond joining the agreement if and only if* $p \geq \overline{p}_H$ *and* $q \geq \overline{q}_L$, *where* $\overline{p}_H \equiv \frac{-BS_{2_H}(\theta_L)}{BS_{2_H}(\theta_H)-BS_{2_H}(\theta_L)}$

In the separating equilibrium, the follower infers that the leader's type must be high after observing a signature and responds joining the treaty, irrespective of its own type. This result

[15] For simplicity, in this section we consider that countries' types are uncorrelated (i.e., independently distributed), whereas section 5 examines how our equilibrium results are affected by the presence of correlated types.

[16] Note that if, in contrast, the leader announces its non-participation in the treaty, the follower's strategy space becomes empty, and hence its posterior beliefs are unconsequential.

[17] We consider that the commitment levels specified in the treaty are above those in the status quo, implying a Pareto improvement.

hence resembles that under complete information, where agreements are only successful when the leader is a high-type country. Intuitively, the low-type leader is not attracted to participate in the treaty since the probability of facing a high-type follower (who will respond joining the agreement in the separating equilibrium) is sufficiently low, $q < \overline{q}_L$. Hence, only the high-type leader signs the treaty, and such signature conveys its type to the follower, thus inducing it to join. The figure below represents the set of pure-strategy equilibria described in Proposition 1, where the separating equilibrium arises when $q < \overline{q}_L$.[18]

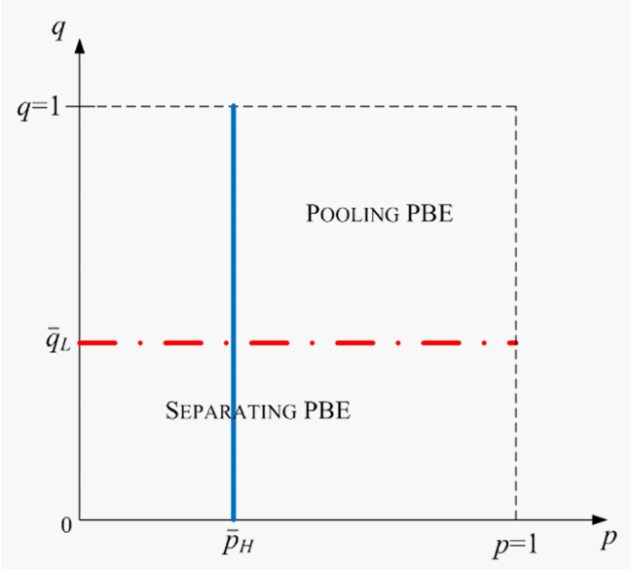

Figure 1. Pure-strategy PBEs.

The introduction of incomplete information, however, allows for the emergence of an equilibrium outcome that could not be sustained under complete information contexts. In particular, a pooling equilibrium is supported where both types of leader sign the agreement and, as a consequence, no information is conveyed to the follower. Hence, this country must decide whether or not to join the treaty based on its expected payoff, which increases in the probability of facing a high-type leader, p. Specifically, the follower signs if p is sufficiently large, $p \geq \overline{p}_H$. Anticipating the follower's signature, both types of leader participate if the follower's type is likely to be high, i.e., $q \geq \overline{q}_L$, and the treaty is successful.[19]

The set of parameter values under which the above separating equilibrium can be sustained depends upon free-riding incentives. In particular, the low-type leader's free-riding incentives are represented by $BS_{1_L}(\theta_H)$, since this expression reflects the leader's benefit from signing an agreement with a high-type follower. When such free-riding incentives decrease, the high-type follower is more attracted to participate in the agreement, expanding the set of priors, q, under which the

[18] Note that cutoffs $\overline{p}_H$ and $\overline{q}_L$ are both positive and lie below 1/2, as described in the proof of Proposition 1.

[19] In our above example, cutoff $\overline{q}_L$ becomes $\overline{q}_L = \frac{NC}{0.69}$ where $NC < 0.39$, implying that $\overline{q}_L \in [0, 0.55]$; while cutoff $\overline{p}_H$ becomes $\overline{p}_H = 0.62$.

separating equilibrium can be sustained, i.e., producing an upward shift in cutoff $\overline{q}_L$ of figure 1.

Let us now examine the region of priors supporting the pooling equilibrium. For a given benefit from signing with a low-type leader, $BS_{2_H}(\theta_L)$, an increase in $BS_{2_H}(\theta_H)$ expands the set of priors under which the follower chooses to sign the agreement in this pooling equilibrium, i.e., produces a leftward shift in cutoff $\overline{p}_H$ in the above figure. Intuitively, the treaty becomes more attractive for the follower when the benefits from signing an agreement with a high-type leader increase, thus inducing the follower to sign.

The previous proposition described the set of equilibria under different conditions on the prior probabilities. However, no equilibrium involving pure strategies exists in the region where $p < \overline{p}_H$ and $q \geq \overline{q}_L$. The following proposition identifies an equilibrium under these parameter conditions in which countries use mixed strategies.

Proposition 2. *In the signaling game where all countries are uninformed about each others' types, a semiseparating equilibrium can be supported when $p < \overline{p}_H$ and $q \geq \overline{q}_L$, where:*

1. *The leader signs with probability $p_L = \frac{p}{1-p}\frac{\overline{p}_H}{1-\overline{p}_H}$ when its type is low, where $p_L \in (0,1)$, and signs the treaty with probability one when its type is high, $p_H = 1$; and*

2. *The follower responds by joining the agreement with probability $r_H(q) \in (0,1)$ when its type is high, and joining the treaty with probability one when its type is low, $r_L = 1$, where its posterior beliefs are $\mu_H(H|S) = \overline{p}_H$, and*

$$r_H(q) \equiv \frac{CUT_{1_L}(\theta_H)}{BPS_{1_L}(\theta_H)} + \frac{1-q}{q}\frac{-BS_{1_L}(\theta_L)}{BPS_{1_L}(\theta_H)}$$

where $CUT_{1_L}(\theta_J) \equiv V_{1_L}(NS;\theta_J) - V_{1_L}(S,NS;\theta_J)$, and $BPS_{1_L}(\theta_J) \equiv V_{1_L}(S,S;\theta_J) - V_{1_L}(S,NS;\theta_J)$.

Therefore, the agreement is signed by both types of leaders and followers with a strictly positive probability. This result, combined with those of Proposition 1, predicts that when countries face uncertainty about each others' types the treaty becomes successful not only when the leader's type is high —as in complete information settings— but also when its type is low. The following figure summarizes the equilibrium predictions of Propositions 1a, 1b and 2, spanning all values of priors $p, q \in [0,1]$. In the next corollary, we examine the comparative statics of players' mixed strategies in the semiseparating equilibrium.

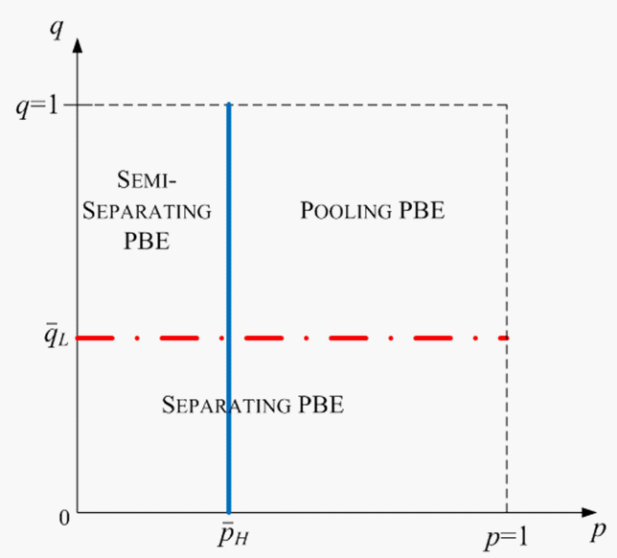

Figure 2. Pure- and mixed-strategy PBEs.

Corollary 1. *The leader's probability of signing the agreement when its type is low, p_L, is: (1) increasing in the probability of the leader being a high type, p; and (2) increasing in the (negative) benefit that the high-type follower obtains from signing a treaty with a low-type leader, $BS_{2_H}(\theta_L)$. Furthermore, the follower's probability of signing the treaty, $r_H(q)$, is: (1) increasing in the costs that the low-type leader suffers from an unsuccessful treaty with a high-type follower, $CUT_{1_L}(\theta_H) \equiv V_{1_L}(NS;\theta_H) - V_{1_L}(S,NS;\theta_H)$; and (2) decreasing in the benefits that the low-type leader obtains from the posterior signature of the treaty by a high-type follower, $BPS_{1_L}(\theta_H) \equiv V_{1_L}(S,S;\theta_H) - V_{1_L}(S,NS;\theta_H)$.*

Let us describe the intuition behind the above corollary. First, an increase in the probability of the leader being a high type raises the follower's incentives to sign the treaty, increasing as a result the low-type leader's probability to participate in the agreement, p_L. Second, an increase in the (negative) benefit that the high-type follower obtains when signing a treaty with a low-type leader reduces the set of beliefs for which the follower is attracted to respond joining the agreement. Hence, in order to be perceived as a high-type country, the low-type leader participates in the treaty with a higher probability, p_L. On the other hand, an increase in the costs that the low-type leader faces when the treaty is unsuccessful makes the signature of the agreement more costly, reducing the likelihood that a signature originates from a low type. As a consequence, the follower is more likely to face a high-type leader, raising the probability $r_H(q)$ with which the follower joins the treaty. In contrast, an increase in the benefit that the low-type leader obtains from the high-type follower's posterior signature of the agreement, $BPS_{1_L}(\theta_H)$, raises the leader's incentive to participate in the treaty. As a result, the likelihood that the follower faces a low-type leader increases, ultimately reducing the follower's probability of joining the treaty.

Information transmission. Let us investigate how the parameters of the model affect the extent of informativeness of the semiseparating equilibrium. A sensible measure is the difference between p_H, the probability with which the high-type leader signs the treaty, and p_L, the probability that the low-type leader participates, i.e., $p_H - p_L$. A larger discrepancy in the probabilities of signature by the two types of leader implies that, when a signature is observed, it is more likely to originate from a high-type leader. On one hand, an increase in the probability of the leader being high, p, raises p_L, reducing the degree of informativeness of the semiseparating equilibrium. Intuitively, a likely high-type leader makes the treaty more attractive for the follower and, as a consequence, also for the low-type leader, who participates in the agreement with a larger probability p_L.[20] Graphically, an increase in p moves our equilibrium predictions, from the semiseparating to the pooling equilibrium where both types of leader sign; as illustrated by arrow (1) in figure 3. On the other hand, a decrease in the probability of the follower being high-type, q, increases $r_H(q)$. As a result, both types of follower respond signing the agreement, converging to their behavior in the separating equilibrium of the game; as represented by arrow (2). Finally, note that the probability with which both types of leader sign the treaty, p_H and p_L, is independent of q. Thus, a change in q does not affect the degree of informativeness of the semiseparating equilibrium, as measured by $p_H - p_L$.

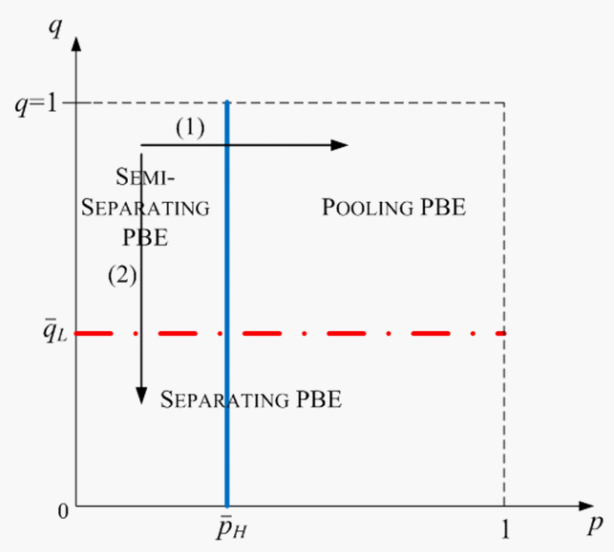

Figure 3. Information transmission.

[20] When the probability of facing a high-type leader is sufficiently large, i.e., $p \geq \overline{p}_H$, the follower and, as a result, the low-type leader sign the treaty (in pure strategies), i.e., $p_L = 1$, behaving as prescribed in the pooling equilibrium of the game.

3.1 Special cases: One-sided uncertainty

Our previous results also provide equilibrium predictions in information settings where only one country is privately informed about its own type, whereas the other country's type is common knowledge. The following corollary investigates the case in which the follower has a long history of compliance (or not compliance) with the content of the treaty, while the leader does not, i.e., thus limiting uncertainty to the leader's type alone.

Corollary 2 (Special case I) *In the signaling game where the leader privately observes its type, while the follower's high type is common knowledge (i.e., $q = 1$), only the pooling equilibrium of Proposition 1b and the semiseparating equilibrium of Proposition 2 can be sustained as PBEs of the game. If, instead, the follower's low type is common knowledge ($q = 0$), only the separating equilibrium of Proposition 1a can be supported as a PBE.*

First, note that this result implies that when countries are informed about the follower's type being high, no separating equilibrium in pure strategies can be sustained, thus limiting the potential of information transmission. As figure 3 above illustrates, when the follower's type is high (i.e., $q = 1$, in the upper dashed horizontal line), only a pooling and a semiseparating PBE can be sustained. Intuitively, the low-type leader is attracted to participate in the treaty under all values of p since it anticipates that the follower will either respond signing —as prescribed in the pooling equilibrium—, or randomize its participation decision, as in the semiseparating equilibrium. If, by contrast, the follower's type is low (i.e., $q = 0$, along the horizontal axis of figure 3), the leader anticipates that the follower will respond joining under all parameter conditions, but subsequently will not implement the content of the treaty. Therefore, only the high-type leader finds beneficial to participate in the treaty with such type of follower since $BS_{1_H}(\theta_L) > 0$. As a consequence, only a separating equilibrium where the high-type leader signs the agreement can be supported.

Let us next investigate the implications of our results under the opposite information context: that arising when the leader's type is common knowledge, whereas the follower privately observes its type; which we refer as special case II. Unlike the information setting examined in Corollary 2, this case illustrates strategic contexts where the country leading the negotiations has a long history of fulfilling the content of the agreements in which it participates or, on the contrary, systematically not living up to its promises. The follower is therefore perfectly informed about the leader's type.

Corollary 3 (Special case II) *In the signaling game where the follower privately observes its type, while the leader's high type is common knowledge (i.e., $p = 1$), only the separating equilibrium of Proposition 1a and the pooling equilibrium of Proposition 1b can be sustained as PBEs of the game. If, instead, the leader's low type is common knowledge ($p = 0$), only the separating equilibrium of Proposition 1a and the semiseparating equilibrium of Proposition 2 can be supported as PBEs.*

First, note that when the leader's type is high (i.e., $p = 1$, in the dashed vertical line of figure 3), both types of followers respond joining a treaty with the leader, since the latter will fully

comply with the content of the agreement. The high-type leader, hence, participates in the treaty. Similarly, when the leader's type is low (i.e., $p = 0$, along the vertical axis of the figure), the follower responds not joining the agreement and, anticipating such a response, the low-type leader does not sign the treaty either.[21]

Finally, equilibrium outcomes under complete information (Lemma 1) are also embodied by our results. In particular, the vertices of the figure depict four possible combinations where: (1) both leader and follower are high type ($p = q = 1$) inducing the leader to sign the treaty and the follower to respond joining; (2) both leader and follower are low type ($p = q = 0$) and the leader does not sign the agreement; (3) the leader is a high type but the follower is not ($p = 1$ and $q = 0$) inducing the leader to sign the treaty and the follower to respond joining; and (4) the leader is a low type but the follower's type is high ($p = 0$ and $q = 1$) inducing the leader to not sign the treaty. Hence, under complete information the agreement is only successful when the leader's type is high (cases 1 and 3), as prescribed in Lemma 1.

4 Welfare comparisons

Let us evaluate the welfare resulting from our previous equilibrium outcomes. We examine the welfare effects of increasing the degree of uncertainty, which is graphically represented in figure 3 by a movement from one of the boundaries to an interior point, where $(p, q) \in (0, 1)$.

Proposition 3. *Equilibrium welfare satisfies the following ranking:*

1. *When the leader's type is high, social welfare in the pooling PBE is weakly larger than under all other equilibrium outcomes,* $SW_{pooling} = SW_{separ} \geq SW_{semisepar}$*, for any follower's type.*

2. *Similarly, when the leader's type is low but the follower's is high, the pooling PBE yields a larger welfare than any of the other equilibrium outcomes if the aggregate benefits from signing the treaty are sufficiently high, that is,*

$$V_{1_L}(S, S; \theta_H) + V_{2_H}(S, S; \theta_L) > V_{1_L}(NS; \theta_H) + V_{2_H}(NS; \theta_L).$$

3. *In contrast, when both countries' types are low, the pooling PBE produces a lower welfare than all other equilibrium outcomes,* $SW_{separ} > SW_{semisepar} > SW_{pooling}$*.*

4. *Finally, under complete information, social welfare coincides with* SW_{separ}*.*

Therefore, when the type of at least one of the countries involved in the negotiations is high, social welfare in the pooling equilibrium, where both types of leader sign and the follower responds

[21] Interestingly, this result applies to both the separating equilibrium —where the low-type leader does not participate, as prescribed in Proposition 1a— and to the semiseparating equilibrium, where the low-type leader randomizes with a probability p_L which collapses to zero when $p = 0$, thus yielding the same equilibrium outcome.

joining, is weakly larger than under any of the other equilibrium outcomes, where either: only one type of leader participates (as in the separating equilibrium), or the leader randomizes its participation decision (as in the semiseparating equilibrium).[22] In contrast, when both countries' types are low, the no signature of the agreement prescribed in the separating equilibrium yields a larger social welfare. If instead the agreement were successful, countries' actions would coincide with those under no treaty but the leader would bear a cost from the unsuccessful treaty, thus entailing a lower social welfare than under the separating equilibrium.

In addition, in complete information contexts countries behave as prescribed in the separating equilibrium. When at least one country's type is high, the introduction of incomplete information yields a social welfare which is not necessarily superior to that under complete information. Specifically, this welfare comparison depends on countries' priors. When priors are relatively symmetric, i.e., $p \geq \overline{p}_H$ and $q \geq \overline{q}_L$, the pooling equilibrium arises, generating a larger welfare than under complete information. By contrast, when priors are relatively asymmetric, i.e., $p < \overline{p}_H$ and $q \geq \overline{q}_L$, the semiseparating equilibrium emerges, entailing a lower social welfare than in complete information contexts.[23] Finally, when both countries' types are low, the social welfare under complete information coincides with that arising in the separating equilibrium, which it is strictly larger than the welfare under the pooling and semiseparating equilibria of the incomplete information game.

5 Correlated types

Consider a setting where countries' types are not independent but, instead, exhibit correlation. When positively correlated, for instance, a country's observation of its high type allows it to infer that the other country is likely to be a high-type country as well. A converse argument applies to types that are negatively correlated. More precisely, let the J-type follower assign a conditional probability $p^J(p)$ to the leader's type being high, and $1 - p^J(p)$ to its type being low. Similarly, the K-type leader assigns a conditional probability $q^K(q)$ to the follower being a high type, and $1 - q^K(q)$ to the follower being low. Note that when priors are uncorrelated (as in previous sections), conditional probabilities satisfy $p^H(p) = p^L(p) = p$ and $q^H(p) = q^L(p) = q$, and therefore, a country's private observation of its type does not provide additional information about the other country's type. In contrast, when types are positively correlated, conditional probabilities satisfy $p^H(p) > p > p^L(p)$ and $q^H(q) > q > q^L(q)$. Positively correlated types describe settings where the countries involved in the negotiations, despite being uncertain about each others' types, share similar institutional contexts, history, etc., which suggests that countries' types are likely to coincide. A converse argument is applicable when priors are negatively correlated, where $p^H(p) < p < p^L(p)$

[22] When the leader's type is low but the follower's is high, the semiseparating equilibrium yields a larger welfare than the separating PBE if the aggregate benefits from signing the treaty are sufficiently high. In such case, the complete welfare ranking is $SW_{pooling} > SW_{semisepar} > SW_{separ}$. For more details, see proof of Proposition 3 in the appendix.

[23] Note that when priors satisfy $p < \overline{p}_H$ and $q < \overline{q}_L$ (or when $p \geq \overline{p}_H$ and $q < \overline{q}_L$), a separating equilibrium arises where social welfare coincides with that under complete information. We hence focus on the regions of prior probabilities for which pooling or semiseparating equilibria emerge.

and $q^H(q) < q < q^L(q)$. The next proposition analyzes how our previous equilibrium predictions are affected by the presence of correlation on players' types.

Proposition 4. *The pooling PBE of Proposition 1b and the semiseparating PBE of Proposition 2 can be supported under a more restrictive set of priors q when types are positively correlated than otherwise. In contrast, the separating PBE of Proposition 1a can be sustained under a larger set of priors q when types are positively correlated than otherwise. The converse result applies when countries' types are negatively correlated.*

The above proposition describes the effects of positive correlation on the pooling equilibrium, where the low-type leader signs under a more restrictive set of priors q. In particular, under uncorrelated types, the low-type leader participates in the treaty only when the probability of facing a high-type follower is sufficiently large. Under positively correlated types, this leader infers that the follower's type is likely to be low as well, and thus becomes more "hesitant" to participate, i.e., signs for a more restricted set of parameter values, ultimately shrinking the set of priors under which the pooling PBE can be sustained. The semiseparating PBE can be similarly supported under more restrictive priors. Intuitively, the low-type leader is less willing to randomize its participation decision under correlated types, given that it is more "certain" that the follower must also be low.

In contrast, the separating equilibrium can be sustained under larger conditions. Specifically, when countries' types are uncorrelated, the low-type leader refrains from signing the treaty if the probability of facing a high-type follower is sufficiently low. When types are correlated, its privately observed low type informs the leader that the follower is likely a low-type country as well, thereby enlarging the set of priors under which the low-type leader does not participate in the treaty, as prescribed in the separating PBE. Hence, the presence of positive correlation facilitates the emergence of informative equilibria, where the leader's signature decision conveys information about its type to the follower, while it hinders the existence of uninformative equilibria (such as the pooling and semiseparating equilibria), whereby information is concealed from the follower.

Finally, when countries' types are negatively correlated, the low-type leader infers that the follower is likely a high-type country, inducing the leader to participate in the agreement under larger parameter conditions, both in the pooling and separating equilibria. Conversely, the low-type leader is attracted to participate in the separating equilibrium under a more restrictive set of priors, thus shrinking the region of parameter values supporting this equilibrium.

Policy implications. From a policy perspective, the results of the paper suggest that negotiations under incomplete information should favor contexts in which one of the countries has a relatively long history of complying with similar agreements. In addition, rather than promoting negotiations between countries with similar domestic constraints and institutions, our equilibrium predictions indicate that international organizations should actually support treaties where countries differ in their political ability to comply with the content of the agreement. Finally, the promotion of treaties where countries' previous compliance history is relatively similar (i.e., symmetric priors) would yield welfare-improving outcomes.

6 Conclusions

Our paper examines bilateral negotiations in a setting where all parties are uninformed about each others' ability to comply with the terms of the agreement. We identify parameter conditions under which either a separating, a pooling or a semiseparating equilibrium exists. The paper also investigates under which conditions information transmission is promoted by changes in the probability that the leader or follower's type is high. Under incomplete information, we demonstrate that the pooling equilibrium yields a larger social welfare than any other equilibrium outcome if at least one country is highly committed with the terms of the treaty. We then examine the welfare properties of our equilibrium results, showing that the presence of incomplete information provides more cases for which the treaty becomes successful, entailing a welfare improvement if, in addition, countries' priors are relatively symmetric. Finally, we extend our analysis to settings where countries' types are positively correlated, finding that the separating equilibrium can be supported under larger conditions, whereas the pooling and semiseparating equilibria are sustained under a more restrictive set of parameter values.

The model describes a context where countries interact only once. However, an enlarged setting could consider a repeated game structure, where signatory countries renegotiate the terms of the agreement or interact in the negotiation of new treaties. In such a context, countries' beliefs not only depend on the signature decisions but also on the compliance history of their cosignatories. Hence, a country's current fulfillment of the agreement would be determined by its political ability to comply, as in our model, and by the future reputational consequences that such compliance entails. In addition, the paper considers that the domestic political situation is unaffected by the signature and posterior implementation of the agreement. However, a different setting could capture the effect of these decisions into a country's political situation (e.g., affecting the likelihood that the incumbent party is reelected) thus modifying the parameter conditions under which our equilibrium outcomes can be supported.

7 Appendix

7.1 *Proof of Lemma 1*

By backward induction, let us first analyze the follower's strategy. When it observes a treaty signed by a high-type leader, both types of follower respond signing since $BS_{2_H}(\theta_H) > 0$ when the follower's type is high and $BS_{2_L}(\theta_H) > 0$ when its type is low. However, if the follower observes a low-type leader signing, it responds not joining the treaty regardless of its type, provided that $BS_{2_H}(\theta_L) < 0$ and $BS_{2_L}(\theta_L) = 0$.

Given the follower's strategy, the high-type leader anticipates that its signature will be responded with a signature, regardless of the follower's type. Hence, it signs the treaty since the benefits from signing satisfy $BS_{1_H}(\theta_H) > BS_{1_H}(\theta_L) > 0$. When the leader's type is low, however,

the leader anticipates that the follower will respond not signing the agreement, regardless of its type, yielding $V_{1_L}(S, NS; \theta_J)$ for $J = \{H, L\}$. Thus, the leader does not participate since by not signing it avoids the cost from an unsuccessful treaty, i.e., $V_{1_L}(NS; \theta_J) > V_{1_L}(S, NS; \theta_J)$. ■

7.2 *Proof of Proposition 1*

Separating equilibrium. Let us first show that the separating strategy profile in which the leader chooses to sign (not sign) the IEA when its type is high (low, respectively) can be supported as a PBE of the signaling game. Under such strategy profile, the J-type follower's beliefs are updated according to Bayes' rule and become $\mu_J(H|S) = 1$ and $\mu_J(H|NS) = 0$ after observing a signature (not signature, respectively) from the leader, where $J = \{H, L\}$. Given these posterior beliefs, the high-type follower signs the agreement since $V_{2_H}(S, S; \theta_H) > V_{2_H}(S, NS; \theta_H)$, and similarly for the low-type follower where $V_{2_L}(S, S; \theta_H) > V_{2_L}(S, NS; \theta_H)$. As a consequence, the high-type leader chooses to participate in the treaty since $V_{1_H}(S, S; \theta_J) > V_{1_H}(NS; \theta_J)$ for all J-type follower, i.e., signing is a strictly dominant strategy for the high-type leader. In addition, the low-type leader does not participate (as prescribed in this separating equilibrium) if

$$q \times V_{1_L}(S, S; \theta_H) + (1 - q) \times V_{1_L}(S, S; \theta_L) < q \times V_{1_L}(NS; \theta_H) + (1 - q) \times V_{1_L}(NS; \theta_L), \text{ or}$$

$$q < \frac{-BS_{1_L}(\theta_L)}{BS_{1_L}(\theta_H) - BS_{1_L}(\theta_L)} \equiv \overline{q}_L$$

where $BS_{1_L}(\theta_J) \equiv V_{1_L}(S, S; \theta_J) - V_{1_L}(NS; \theta_J)$ denotes the low-type leader's benefit of signing an agreement with a J-type follower. Note that cutoff $\overline{q}_L > 0$ since $BS_{1_L}(\theta_L) < 0$ and $\overline{q}_L < \frac{1}{2}$ given that $BS_{1_L}(\theta_H) > 0$ and $BS_{1_L}(\theta_H) > BS_{1_L}(\theta_L)$. Hence, the low-type leader does not participate if $q < \overline{q}_L$. Therefore, the above separating strategy profile can be sustained as a PBE of the game if $q < \overline{q}_L$.

Pooling equilibrium with signature. Let us next demonstrate that the pooling strategy profile, in which the leader signs the treaty regardless of its type, can be part of a PBE under certain conditions. In this strategy, the J-type follower's posterior beliefs cannot be updated and thus coincide with its priors, i.e., $\mu_J(H|S) = p$ and $\mu_J(L|S) = 1 - p$. (Note that off-the-equilibrium beliefs do not play a role in this pooling equilibrium. In particular, after observing the off-the-equilibrium action of "no signature" the follower has, by definition, an empty action space. Therefore, off-the-equilibrium beliefs cannot affect the follower's response and, as a consequence, do not affect the leader's decision either.)

Given these beliefs, the high-type follower chooses to not sign the agreement if

$$p \times V_{2_H}(S, S; \theta_H) + (1 - p) \times V_{2_H}(S, S; \theta_L) < p \times V_{2_H}(S, NS; \theta_H) + (1 - p) \times V_{2_H}(S, NS; \theta_L), \text{ or}$$

$$p < \frac{-BS_{2_H}(\theta_L)}{BS_{2_H}(\theta_H) - BS_{2_H}(\theta_L)} \equiv \overline{p}_H$$

where $BS_{2_H}(\theta_K) \equiv V_{2_H}(S, S; \theta_K) - V_{2_H}(S, NS; \theta_K)$ denotes the high-type follower's benefit from

signing an agreement with a K-type leader. Note that $\overline{p}_H > 0$ since the follower's payoffs satisfy $BS_{2_H}(\theta_L) < 0$ and $BS_{2_H}(\theta_H) > BS_{2_H}(\theta_L)$ by definition. In addition, $\overline{p}_H < \frac{1}{2}$ given that $BS_{2_H}(\theta_H) - BS_{2_H}(\theta_L) > -BS_{2_H}(\theta_L)$ or $BS_{2_H}(\theta_H) > 0$. Hence, when $p < \overline{p}_H$, the high-type follower does not sign the treaty, and signs otherwise. Similarly, the low-type follower does not sign the agreement if

$$p \times V_{2_L}(S, S; \theta_H) + (1-p) \times V_{2_L}(S, S; \theta_L) < p \times V_{2_L}(S, NS; \theta_H) + (1-p) \times V_{2_L}(S, NS; \theta_L), \text{ or}$$

$$p < \frac{-BS_{2_L}(\theta_L)}{BS_{2_L}(\theta_H) - BS_{2_L}(\theta_L)} \equiv \overline{p}_L$$

where $BS_{2_L}(\theta_K) \equiv V_{2_L}(S, S; \theta_K) - V_{2_L}(S, NS; \theta_K)$ denotes the low-type follower's benefit from signing an agreement with a K-type leader. Since $BS_{2_L}(\theta_L) = 0$ by definition, $\overline{p}_L = 0$, implying that the low-type follower signs for all parameter values. Let us next analyze equilibrium strategies for the leader.

1. *High priors, $p \geq \overline{p}_H$.* In this case both types of follower respond signing the treaty. The high-type leader participates in the agreement since $V_{1_H}(S, S; \theta_K) > V_{1_H}(NS; \theta_K)$ for all follower K. However, the low-type leader signs if

$$qV_{1_L}(S, S; \theta_H) + (1-q)V_{1_L}(S, S; \theta_L) \geq qV_{1_L}(NS; \theta_H) + (1-q)V_{1_L}(NS; \theta_L), \text{ or}$$

$$q \geq \frac{-BS_{1_L}(\theta_L)}{BS_{1_L}(\theta_H) - BS_{1_L}(\theta_L)} \equiv \overline{q}_L$$

   where $\overline{q}_L \in (0, 1)$ from our above discussion in the separating equilibrium. Therefore, the pooling strategy profile in which both types of leader sign the agreement can be sustained if $p \geq \overline{p}_H$ and $q \geq \overline{q}_L$.

2. *Low priors, $p < \overline{p}_H$.* In this case the high-type follower responds not participating in the treaty while the low-type follower signs for all parameter values. The high-type leader participates in the agreement for all priors q since $V_{1_H}(S, NS; \theta_H) > V_{1_H}(NS; \theta_H)$ when facing a high-type follower and $V_{1_H}(S, S; \theta_L) > V_{1_H}(NS; \theta_L)$ when facing a low-type follower. Regarding the low-type leader, he signs the treaty if

$$qV_{1_L}(S, NS; \theta_H) + (1-q)V_{1_L}(S, S; \theta_L) \geq qV_{1_L}(NS; \theta_H) + (1-q)V_{1_L}(NS; \theta_L), \text{ or}$$

$$q \geq \frac{-BS_{1_L}(\theta_L)}{[V_{1_L}(S, NS; \theta_H) - V_{1_L}(NS; \theta_H)] - BS_{1_L}(\theta_L)} \equiv \overline{q}'_L$$

   where $BS_{1_L}(\theta_L) < 0$ by definition. In addition, $V_{1_L}(S, NS; \theta_H) < V_{1_L}(NS; \theta_H)$ since the agreement is not successful. Hence, cutoff $\overline{q}'_L > 1$ and the low-type leader does not sign the agreement for any prior q. Therefore, the pooling strategy profile in which both types of leader sign the treaty cannot be sustained as PBE when $p < \overline{p}_H$.

Pooling equilibrium with no signature. Finally, let us show that the pooling strategy

profile where the leader does not sign the agreement regardless of its type cannot be sustained as part of a PBE. First, note that the follower's posterior beliefs cannot be updated using Bayes' rule, and hence must be arbitrarily specified, $\mu_J(H|S) \in [0,1]$ for any J-type follower, where $J = \{H, L\}$. Given these beliefs, the high-type follower signs the agreement if and only if

$$\begin{aligned} & \mu_H(H|S) \times V_{2_H}(S,S;\theta_H) + (1-\mu_H(H|S)) \times V_{2_H}(S,S;\theta_L) \\ \geq \; & \mu_H(H|S) \times V_{2_H}(S,NS;\theta_H) + (1-\mu_H(H|S)) \times V_{2_H}(S,NS;\theta_L), \text{ or} \end{aligned}$$

$$\mu_H(H|S) \geq \frac{-BS_{2_H}(\theta_L)}{BS_{2_H}(\theta_H) - BS_{2_H}(\theta_L)} \equiv \overline{p}_H$$

where $\overline{p}_H \in (0,1)$ from our above discussion. Regarding the low-type follower, it participates in the treaty if

$$\begin{aligned} & \mu_L(H|S) \times V_{2_L}(S,S;\theta_H) + (1-\mu_L(H|S)) \times V_{2_L}(S,S;\theta_L) \\ \geq \; & \mu_L(H|S) \times V_{2_L}(S,NS;\theta_H) + (1-\mu_L(H|S)) \times V_{2_L}(S,NS;\theta_L), \text{ or} \end{aligned}$$

$$\mu_L(H|S) \geq \frac{-BS_{2_L}(\theta_L)}{BS_{2_L}(\theta_H) - BS_{2_L}(\theta_L)} \equiv \overline{p}_L$$

where $\overline{p}_L = 0$ from our above discussion. Hence, the low-type follower signs the agreement for all off-the-equilibrium beliefs $\mu_L(H|S)$. Let us now analyze the leader's strategy.

1. If $\mu_H(H|S) \geq \overline{p}_H$, both the high and low-type follower respond signing. If the leader is a high type country, it signs the agreement since $V_{1_H}(S,S;\theta_J) > V_{1_H}(NS;\theta_J)$ for any J-type of follower. Hence, the high-type leader signs the treaty under all priors q and the pooling strategy profile in which no leader signs cannot be supported as PBE for the case in which off-the-equilibrium beliefs satisfy $\mu_H(H|S) \geq \overline{p}_H$.

2. If $\mu_H(H|S) < \overline{p}_H$, the high-type follower responds not signing whereas the low-type responds signing. If the leader is a high type country, it participates in the agreement since $V_{1_H}(S,NS;\theta_H) > V_{1_H}(NS;\theta_H)$ when facing a high-type follower and $V_{1_H}(S,S;\theta_L) > V_{1_H}(NS;\theta_L)$ when facing a low-type follower. The pooling strategy profile in which both types of leader do not sign the treaty cannot sustained as a PBE of the game when off-the-equilibrium beliefs satisfy $\mu_H(H|S) < \overline{p}_H$.

Intuitive Criterion. Let us apply the Cho and Kreps' (1987) Intuitive Criterion for the pooling PBE where $p \geq \overline{p}_H$ and $q \geq \overline{q}_L$. We first check if a deviation towards "not sign" is equilibrium dominated for either type of leader. When the leader is a high type country, the highest payoff that it obtains by deviating towards "not sign" is $qV_{1_H}(NS;\theta_H) + (1-q)V_{1_H}(NS;\theta_L)$, which does exceed its equilibrium payoff, $qV_{1_H}(S,S;\theta_H) + (1-q)V_{1_H}(S,S;\theta_L)$, since $V_{1_H}(S,S;\theta_J) > V_{1_H}(NS;\theta_J)$ for any J-type follower. Regarding the low-type leader, the highest payoff that it can obtain by deviating is $qV_{1_L}(NS;\theta_H) + (1-q)V_{1_L}(NS;\theta_L)$ which exceeds its equilibrium payoff of

$qV_{1_L}(S,S;\theta_H)+(1-q)V_{1_L}(S,S;\theta_L)$ if

$$qV_{1_L}(S,S;\theta_H)+(1-q)V_{1_L}(S,S;\theta_L)<qV_{1_L}(NS;\theta_H)+(1-q)V_{1_L}(NS;\theta_L), \text{ or } q<\overline{q}_L$$

where cutoff $\overline{q}_L\in(0,1)$ from our above discussion in the separating equilibrium. Hence, the low-type leader deviates towards "not sign" if $q<\overline{q}_L$. Therefore, the pooling equilibrium in which both types of leader do not participate in the treaty (supported under $p\geq\overline{p}_H$ and $q\geq\overline{q}_L$) survives the Cho and Kreps' (1987) Intuitive Criterion. ■

7.3 *Proof of Proposition 2*

Let us first analyze the strategy for the high-type follower. (The low-type follower signs the agreement for all priors p, and therefore it does not modify its signature decision based on the information inferred from the leader's randomization).The high-type follower must be mixing. Otherwise, the leader could anticipate its action and play pure strategies as in any of the strategy profiles described in proposition 2, which are not PBE of the signaling game when $p<\overline{p}_H$ and $q\geq\overline{q}_L$. Hence, the high-type follower must be indifferent between signing and not signing the treaty, that is,

$$\begin{aligned}&\mu_{2_H}(H|S)\times V_{2_H}(S,S;\theta_H)+(1-\mu_{2_H}(H|S))\times V_{2_H}(S,S;\theta_L)\\ =\;&\mu_{2_H}(H|S)\times V_{2_H}(S,NS;\theta_H)+(1-\mu_{2_H}(H|S))\times V_{2_H}(S,NS;\theta_L),\end{aligned}$$

or $\mu_{2_H}(H|S)=\overline{p}_H$. We can next use the follower's posterior beliefs in order to find the probability with which the leader randomizes when its type is low, p_L, by using Bayes' rule.

$$\mu_H(H|S)=\overline{p}_H=\frac{(1-p)\times p_L}{((1-p)\times p_L)+(p\times p_H)}$$

where $p_H=1$. Solving for p_L, we obtain $p_L=\frac{p}{1-p}\frac{\overline{p}_H}{1-\overline{p}_H}$, which is positive, $p_L>0$, since $p,\overline{p}_H\in(0,1)$. In addition, note that $p_L<1$ holds for all priors p satisfying $p<1-\overline{p}_H$. Since $|BS_{2_H}(\theta_H)|>|BS_{2_H}(\theta_L)|$ by definition, cutoff $\overline{p}_H$ satisfies $\overline{p}_H<1/2$, which implies $p<1-\overline{p}_H$ holds for all priors, because $p<\overline{p}_H$ by assumption. Note that probability p_L increases in p.

Finally, if the low-type leader mixes with probability $p_L\in(0,1)$, it must be that the high-type follower makes this leader indifferent between signing and not signing the agreement (the low-type follower responds by signing under all parameter conditions). Using $r_H(q)$ to denote the probability with which the high-type follower mixes between signing and not signing the treaty, the low-type leader is indifferent if

$$\begin{aligned}&q\left[r_H(q)\times V_{1_L}(S,S;\theta_H)+(1-r_H(q))\times V_{1_L}(S,NS;\theta_H)\right]+(1-q)V_{1_L}(S,S;\theta_L)\\ =\;&qV_{1_L}(NS;\theta_H)+(1-q)V_{1_L}(NS;\theta_L), \text{ or}\end{aligned}$$

$$r_H(q)=\frac{CUT_{1_L}(\theta_H)}{BPS_{1_L}(\theta_H)}+\frac{1-q}{q}\frac{-BS_{1_L}(\theta_L)}{BPS_{1_L}(\theta_H)}$$

where $BPS_{1_L}(\theta_J) \equiv V_{1_L}(S,S;\theta_J) - V_{1_L}(S,NS;\theta_J)$ denotes the benefits that the low-type leader obtains from the posterior signature of the agreement by the J-type follower. On the other hand, $CUT_{1_L}(\theta_J) \equiv V_{1_L}(NS;\theta_J) - V_{1_L}(S,NS;\theta_J)$ represents the low-type leader's cost from an unsuccessful treaty. It is easy to show that this leader's benefit from signing the treaty, $BS_{1_L}(\theta_J)$, can therefore be expressed as the sum of the above two benefits, i.e., $BS_{1_L}(\theta_J) = BPS_{1_L}(\theta_J) - CUT_{1_L}(\theta_J)$, where $J = \{H, L\}$. In addition, note that the probability $r_H(q)$ satisfies $r_H(q) \in (0,1)$. Indeed, $CUT_{1_L}(\theta_H) + \frac{1-q}{q}[-BS_{1_L}(\theta_L)] < BPS_{1_L}(\theta_H)$ implies $\frac{1-q}{q} < \frac{BS_{1_L}(\theta_H)}{-BS_{1_L}(\theta_L)}$, which can be rearranged as $\overline{q}_L \equiv \frac{-BS_{1_L}(\theta_L)}{BS_{1_L}(\theta_H) - BS_{1_L}(\theta_L)} < q$, which holds by assumption. Finally, note that probability $r_H(q)$ is decreasing in q. ■

7.4 *Proof of Corollary 2*

Follower's type is high. When $q = 1$, then condition $q < \overline{q}_L$ does not hold, and thus the separating equilibrium of Proposition 1a cannot be sustained. The pooling equilibrium of Proposition 1b, however, can be supported if the prior probability p satisfies condition $p \geq \overline{p}_H$. The semiseparating equilibrium of Proposition 2 can be sustained (since $q = 1 \geq \overline{q}_L$ holds) if, in addition, prior p satisfies $p < \overline{p}_H$. In such case, the low-type (high-type) leader signs with a probability $p_L = \frac{p}{1-p}\frac{\overline{p}_H}{1-\overline{p}_H}$ ($p_H = 1$, respectively), which is unaffected by the value of q, given that $\overline{p}_H$ is independent on q. The high-type follower responds joining the treaty with probability $r_H(q)$, which decreases in q, becoming $r_H(1) = \frac{CUT_{1_L}(\theta_H)}{BPS_{1_L}(\theta_H)}$ when $q = 1$; whereas the low-type follower responds signing with probability one, irrespective of the value of q.

Follower's type is low. When $q = 0$, then condition $q < \overline{q}_L$ is satisfied, and the separating equilibrium of Proposition 1a can be sustained. The pooling equilibrium of Proposition 1b, however, cannot be supported since $q \geq \overline{q}_L$ is not satisfied when $q = 0$. The semiseparating equilibrium of Proposition 2 cannot be sustained either since condition $q \geq \overline{q}_L$ does not hold. ■

7.5 *Proof of Corollary 3*

Leader's type is high. When $p = 1$, then condition $q < \overline{q}_L$ holds, and thus the separating equilibrium of Proposition 1a can be sustained. Similarly, the pooling equilibrium of Proposition 1b can be supported since condition $p \geq \overline{p}_H$ holds when $p = 1$. The semiseparating equilibrium of Proposition 2, however, cannot be sustained since condition $p < \overline{p}_H$ is not satisfied when $p = 1$.

Follower's type is low. When $p = 0$, condition $q < \overline{q}_L$ is satisfied, and therefore the separating equilibrium of Proposition 1a can be sustained. The pooling equilibrium of Proposition 1b, however, cannot be supported since condition $p \geq \overline{p}_H$ is not satisfied when $p = 0$. Finally, the semiseparating equilibrium of Proposition 2 can be sustained since condition $p < \overline{p}_H$ holds when $p = 0$. In this setting, the low-type leader signs with a probability $p_L = \frac{p}{1-p}\frac{\overline{p}_H}{1-\overline{p}_H}$, which collapses to $p_L = 0$ when $p = 0$, i.e., the low-type leader does not participate in the treaty in pure strategies. The high-type leader, by contrast, signs with probability $p_H = 1$, which is independent on p. Let us now analyze the follower. First, the high-type follower responds joining the treaty with probability

$r_H(q) = \frac{CUT_{1_L}(\theta_H)}{BPS_{1_L}(\theta_H)} + \frac{1-q}{q}\frac{-BS_{1_L}(\theta_L)}{BPS_{1_L}(\theta_H)}$ which does not depend upon the prior p. Similarly, the low-type follower responds singing with probability $r_L = 1$, irrespective of the value of p. ■

7.6 *Proof of Proposition 3*

Separating PBE. When the leader's type is high, it participates in the agreement and the follower responds joining, yielding a social welfare (summing up the equilibrium payoffs of leader and follower) of $V_{1_H}(S, S; \theta_J) + V_{2_J}(S, S; \theta_H)$ where $J = \{H, L\}$ denotes the follower's type. If, in contrast, the leader's type is low, the leader does not sign the treaty, entailing a welfare of $V_{1_L}(NS; \theta_J) + V_{2_J}(NS; \theta_L)$.

Pooling PBE. The pooling PBE yields a social welfare of $V_{1_K}(S, S; \theta_J) + V_{2_J}(S, S; \theta_K)$, where $K = \{H, L\}$ denotes the leader's type and $J = \{H, L\}$ represents the follower's type, which entails the same social welfare as in the separating PBE when the leader's type is high, i.e., $SW_{separ} = SW_{pooling}$, for any follower's type J. However, when the leader's type is low, the separating equilibrium prescribes that this leader does not participate in the treaty, yielding a social welfare of $V_{1_L}(NS; \theta_J) + V_{2_J}(NS; \theta_L)$, which lies weakly below that under the pooling PBE if

$$V_{1_L}(NS; \theta_J) + V_{2_J}(NS; \theta_L) \leq V_{1_L}(S, S; \theta_J) + V_{2_J}(S, S; \theta_L)$$

or alternatively, $-BS_{2_J}(\theta_L) \leq BS_{1_L}(\theta_J)$, where $BS_{1_L}(\theta_J) \equiv V_{1_L}(S, S; \theta_J) - V_{1_L}(NS; \theta_J)$ denotes the low-type leader's benefit from signing an agreement with a J-type follower, and conversely $BS_{2_L}(\theta_J) \equiv V_{2_J}(S, S; \theta_L) - V_{2_J}(S, NS; \theta_L)$ represents the J-type follower's benefit from signing a treaty with a low-type leader, since $V_{2_J}(NS; \theta_L) = V_{2_J}(S, NS; \theta_L)$. Let us separately analyze the cases in which the follower's type is high and low.

1. If the leader's type is low while that of the follower is high, i.e., $J = H$, then $BS_{1_L}(\theta_H) > 0$ for the leader and $BS_{2_H}(\theta_L) < 0$ for the follower and, as a consequence, condition $-BS_{2_H}(\theta_L) \leq BS_{1_L}(\theta_H)$ holds if

$$V_{1_L}(S, S; \theta_H) + V_{2_H}(S, S; \theta_L) > V_{1_L}(NS; \theta_H) + V_{2_H}(NS; \theta_L)$$

   which represents that the aggregated benefits from signing the treaty are sufficiently high, implying that $SW_{separ} < SW_{pooling}$. Otherwise, $-BS_{2_H}(\theta_L) > BS_{1_L}(\theta_H)$, and the separating equilibrium yields a larger social welfare than the pooling equilibrium.

2. When the leader's and follower's type are low, i.e., $J = L$, the benefits from signing an agreement between two low-type countries are $BS_{1_L}(\theta_L) < 0$ for the leader and $BS_{2_L}(\theta_L) = 0$ for the follower. Therefore, when both leader and follower are low types, condition $-BS_{2_J}(\theta_L) \leq BS_{1_L}(\theta_J)$ does not hold, and hence social welfare in the separating equilibrium is larger than under the pooling equilibrium, i.e., $SW_{separ} > SW_{pooling}$.

Semiseparating PBE. Let us now evaluate social welfare in the semiseparating equilibrium.

1. When both the leader and follower's types are high, the former participates with probability one while the latter randomizes according to probability $r_H(q)$. Hence, social welfare becomes

$$r_H(q) \times [V_{1_H}(S,S;\theta_H) + V_{2_H}(S,S;\theta_H)] + (1 - r_H(q)) \times [V_{1_H}(S,NS;\theta_H) + V_{2_H}(S,NS;\theta_H)]$$

   where the first term in brackets, $V_{1_H}(S,S;\theta_H) + V_{2_H}(S,S;\theta_H)$, is larger than the second, $V_{1_H}(S,NS;\theta_H) + V_{2_H}(S,NS;\theta_H)$, since $V_{1_H}(S,S;\theta_H) > V_{1_H}(S,NS;\theta_H)$ for the leader and similarly $V_{2_H}(S,S;\theta_H) > V_{2_H}(S,NS;\theta_H)$ for the follower, since $BS_{2_H}(\theta_H) > 0$. Furthermore, recall that social welfare under the pooling PBE is $V_{1_H}(S,S;\theta_H) + V_{2_H}(S,S;\theta_H)$. Hence, welfare in the semiseparating equilibrium is a linear combination between the social welfare in the pooling PBE and a smaller number, thereby yielding a lower welfare than under the pooling PBE. Combining this result with that from the pooling PBE, we obtain that $SW_{separ} = SW_{pooling} > SW_{semisepar}$.

2. When the leader's type is high but the follower's is low, this equilibrium prescribes that both countries sign the agreement with probability one, thus yielding a social welfare of $V_{1_H}(S,S;\theta_L) + V_{2_L}(S,S;\theta_H)$, which coincides with that under the pooling equilibrium of Proposition 1b. Combining this result with that from the pooling PBE, we obtain that $SW_{separ} = SW_{pooling} = SW_{semisepar}$.

3. Finally, when leader's type is low and that of the follower is high, both countries randomize their participation decision, yielding social welfare of

$$\begin{aligned} &p_L \left[r_H(q) \left[V_{1_L}(S,S;\theta_H) + V_{2_H}(S,S;\theta_L) \right] + (1 - r_H(q)) \left[V_{1_L}(S,NS;\theta_H) + V_{2_H}(S,NS;\theta_L) \right] \right] \\ &+ (1 - p_L) \left[V_{1_L}(NS;\theta_H) + V_{2_H}(NS;\theta_L) \right]. \end{aligned}$$

   whereas social welfare under the pooling PBE is $V_{1_L}(S,S;\theta_H) + V_{2_H}(S,S;\theta_L)$, which we denote as A. Hence, the social welfare in the semiseparating equilibrium is lower than under pooling PBE if

$$\begin{aligned} &p_L \left[r_H(q) A + (1 - r_H(q)) \left[V_{1_L}(S,NS;\theta_H) + V_{2_H}(S,NS;\theta_L) \right] \right] \\ &+ (1 - p_L) \left[V_{1_L}(NS;\theta_H) + V_{2_H}(NS;\theta_L) \right] < A \end{aligned}$$

   rearranging, using the property that $V_{2_H}(S,NS;\theta_L) = V_{2_H}(NS;\theta_L) = X$, and solving for the payoff A, we obtain

$$\overline{A} \equiv \frac{-p_L CUT_{1_L}(\theta_H) + B - p_L r_H(q) \left[V_{1_L}(S,NS;\theta_H) + X \right]}{1 - p_L r_H(q)} < A$$

   where $CUT_{1_L}(\theta_H) \equiv V_{1_L}(NS;\theta_H) - V_{1_L}(S,NS;\theta_H)$, and $B \equiv V_{1_L}(NS;\theta_H) + V_{2_H}(NS;\theta_L)$. Cutoff $\overline{A}$ lies below B, which implies that $A > B$ is a sufficient condition for $A > \overline{A}$, entailing

that welfare in the pooling equilibrium exceeds that in the semiseparating equilibrium. In particular, $\overline{A} < B$ since

$$-p_L CUT_{1_L}(\theta_H) + B - p_L r_H(q) \left[V_{1_L}(S, NS; \theta_H) + X\right] < B \left[1 - p_L r_H(q)\right]$$

which implies

$$\frac{CUT_{1_L}(\theta_H)}{r_H(q)} + \left[V_{1_L}(S, NS; \theta_H) + X\right] > B$$

Using $B \equiv V_{1_L}(NS; \theta_H) + V_{2_H}(NS; \theta_L) = V_{1_L}(NS; \theta_H) + X$, since $V_{2_H}(NS; \theta_L) = X$, the above inequality can be simplified to $CUT_{1_L}(\theta_H) > r_H(q) CUT_{1_L}(\theta_H)$, which holds by assumption given that $r_H(q) \in (0, 1)$. Concluding, condition $A > B$, or alternatively,

$$V_{1_L}(S, S; \theta_H) + V_{2_H}(S, S; \theta_L) > V_{1_L}(NS; \theta_H) + V_{2_H}(NS; \theta_L)$$

guarantees that welfare in the pooling equilibrium exceeds that in the semiseparating equilibrium.

- Finally, in order to obtain a complete welfare ranking, let us now compare welfare under the separating and semiseparating equilibrium when the leader's type is low and the follower's is high. In the separating PBE, social welfare is $V_{1_L}(NS; \theta_H) + V_{2_H}(NS; \theta_L) \equiv B$. The semiseparating PBE yields a welfare of

  $$p_L \left[r_H(q) A + (1 - r_H(q)) \left[V_{1_L}(S, NS; \theta_H) + X\right]\right] + (1 - p_L) \left[V_{1_L}(NS; \theta_H) + X\right],$$

  which is larger than in the separating equilibrium, B, if

  $$A > \frac{p_L CUT_{1_L}(\theta_H) - (1 - p_L) V_{1_L}(S, NS; \theta_H) + r_H(q) \left[V_{1_L}(S, NS; \theta_H) + X\right]}{p_L r_H(q)} \equiv \widetilde{A}$$

  Hence, if $A > \widetilde{A}$ (i.e., the benefits from signing the agreement are sufficiently high) we obtain the complete welfare ranking $SW_{pooling} > SW_{semisepar} > SW_{separ}$. If, instead, $A \leq \widetilde{A}$, the welfare ranking becomes $SW_{pooling} > SW_{separ} > SW_{semisepar}$.

4. When, in contrast, both the leader and follower's types are low, the former randomizes according to a probability $p_L = \frac{p}{1-p} \frac{\overline{p}_H}{1-\overline{p}_H}$, whereas the follower responds joining the treaty with probability one. Therefore, social welfare in this case is

   $$p_L \times \left[V_{1_L}(S, S; \theta_L) + V_{2_L}(S, S; \theta_L)\right] + (1 - p_L) \times \left[V_{1_L}(NS; \theta_L) + V_{2_L}(NS; \theta_L)\right].$$

   which can be alternatively expressed as

   $$p_L \times [V_{1_L}(S, S; \theta_L) - V_{1_L}(NS; \theta_L)] + p_L \times [V_{2_L}(S, S; \theta_L) - V_{2_L}(NS; \theta_L)] + V_{1_L}(NS; \theta_L) + V_{2_L}(NS; \theta_L).$$

where the first term in brackets is negative since $V_{1_L}(S,S;\theta_L) < V_{1_L}(NS;\theta_L)$ for the leader, given that $BS_{1_L}(\theta_L) < 0$ by assumption. The second term is zero because $V_{2_L}(S,S;\theta_L) = V_{2_L}(NS;\theta_L)$ for the follower, since $BS_{2_L}(\theta_L) = 0$ given that $V_{2_L}(NS;\theta_L) = V_{2_L}(S,NS;\theta_L)$. Therefore, welfare under the semiseparating equilibrium becomes

$$p_L \times [V_{1_L}(S,S;\theta_L) - V_{1_L}(NS;\theta_L)] + V_{1_L}(NS;\theta_L) + V_{2_L}(NS;\theta_L).$$

which lies above the welfare under the pooling PBE, $V_{1_L}(S,S;\theta_L) + V_{2_L}(S,S;\theta_L)$, if

$$p_L \times [V_{1_L}(S,S;\theta_L) - V_{1_L}(NS;\theta_L)] + V_{1_L}(NS;\theta_L) + V_{2_L}(NS;\theta_L) > V_{1_L}(S,S;\theta_L) + V_{2_L}(S,S;\theta_L)$$

and since $V_{2_L}(S,S;\theta_L) = V_{2_L}(NS;\theta_L)$ for the follower, the above condition can be expressed as

$$p_L \times [V_{1_L}(S,S;\theta_L) - V_{1_L}(NS;\theta_L)] + V_{1_L}(NS;\theta_L) > V_{1_L}(S,S;\theta_L)$$

and rearranging, we obtain $V_{1_L}(NS;\theta_L) > V_{1_L}(S,S;\theta_L)$, a condition that holds by assumption.

- Finally, in order to obtain a complete welfare ranking, let us now compare social welfare under the separating and semiseparating equilibrium. In the separating PBE, social welfare is $V_{1_L}(NS;\theta_L) + V_{2_L}(NS;\theta_L)$; whereas in the semiseparating PBE social welfare is

  $$p_L \times [V_{1_L}(S,S;\theta_L) - V_{1_L}(NS;\theta_L)] + V_{1_L}(NS;\theta_L) + V_{2_L}(NS;\theta_L).$$

  which lies below the welfare under the separating PBE, $V_{1_L}(NS;\theta_L) + V_{2_L}(NS;\theta_L)$, if

  $$p_L \times [V_{1_L}(S,S;\theta_L) - V_{1_L}(NS;\theta_L)] + V_{1_L}(NS;\theta_L) + V_{2_L}(NS;\theta_L) < V_{1_L}(NS;\theta_L) + V_{2_L}(NS;\theta_L)$$

  and rearranging $p_L\,[V_{1_L}(S,S;\theta_L) - V_{1_L}(NS;\theta_L)] < 0$, which holds by assumption given that $V_{1_L}(S,S;\theta_L) < V_{1_L}(NS;\theta_L)$ when both countries' types are low.
- Therefore, social welfare in the semiseparating PBE is lower than in the separating equilibrium, yielding a complete ranking of $SW_{separ} > SW_{semisepar} > SW_{pooling}$.

Complete information. Let us first compare social welfare when both countries' types are high. Under complete information, social welfare is $SW^{HH}_{complete} \equiv V_{1_H}(S,S;\theta_H) + V_{2_H}(S,S;\theta_H)$, which coincides with equilibrium welfare under the pooling and separating PBE. Regarding the case where the leader's type is high but the follower's is low, $SW^{HL}_{complete} \equiv V_{1_H}(S,S;\theta_L) + V_{2_L}(S,S;\theta_H)$ also coincides with that under the pooling and separating equilibrium. When the leader's type is low but the follower's is high, $SW^{LH}_{complete} \equiv V_{1_L}(NS;\theta_H) + V_{2_H}(NS;\theta_L)$, which coincides with the social welfare under the separating equilibrium, and hence, lies weakly below that under the pooling PBE if

$$V_{1_L}(NS;\theta_J) + V_{2_J}(NS;\theta_L) \leq V_{1_L}(S,S;\theta_J) + V_{2_J}(S,S;\theta_L).$$

Finally, when both countries' types are low, $SW^{LL}_{complete} \equiv V_{1_L}(NS;\theta_L) + V_{2_L}(NS;\theta_L)$, which coincides with social welfare under the separating equilibrium, and therefore, lies above that under the pooling PBE. ■

7.7 *Proof of Proposition 4*

Pooling PBE. When countries' types are uncorrelated, the pooling PBE can be sustained if the high-type follower's priors p satisfy $p \geq \overline{p}_H$. When types are positively correlated, the high-type follower's conditional probability that the leader's type is also high becomes

$$prob(l = H|f = H) = \frac{prob(f = H|l = H) \times prob(l = H)}{prob(f = H)} = \frac{q^H(q)}{q}p \equiv p^H(p)$$

where l (f) denotes the leader (follower, respectively). In addition, given positive correlation, $q^H(q) > q > q^L(q)$, which implies $p^H(p) > p$; as depicted in figure A1 below. In particular, note that for the set of priors p under which the pooling PBE emerges under uncorrelated types, i.e., $p \geq \overline{p}_H$, condition $p^H(p) \geq \overline{p}_H$ holds. Therefore, condition $p^H(p) \geq \overline{p}_H$ can be sustained under a larger set of priors p than condition $p \geq \overline{p}_H$, and the pooling PBE can hence be supported under a larger set of priors p when countries' types are positively correlated than otherwise.

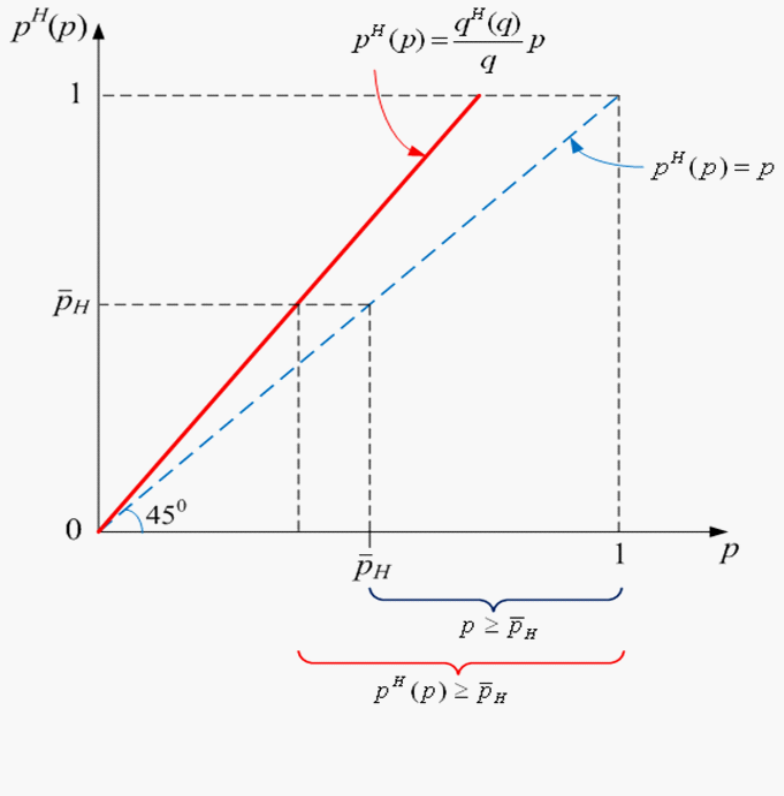

Figure A1

Let us now examine the leader's incentives. Specifically, the low-type leader participates in the treaty if $q \geq \overline{q}_L$ under uncorrelated types. When types are correlated, the low-type leader constructs the conditional probability that the follower is a high type, as follows,

$$prob(f = H|l = L) = \frac{prob(l = L|f = H) \times prob(f = H)}{prob(l = L)} = \frac{1 - p^H(p)}{1 - p} q \equiv q^L(q)$$

and, by positive correlation, $p^H(p) > p > p^L(p)$, implying that ratio $\frac{1-p^H(p)}{1-p}$ satisfies $\frac{1-p^H(p)}{1-p} < 1$. Hence, $q^L(q) < q$; as illustrated in figure A2 below. Furthermore, note that condition $q \geq \overline{q}_L$ can be sustained under a larger set of priors q than condition $q^L(q) \geq \overline{q}_L$. Hence, the pooling PBE can be supported under a more restricted set of priors q when types are correlated than otherwise. When countries' types are negatively correlated, $q^H(q) < q < q^L(q)$, inducing the low-type leader to sign the agreement under a larger set of priors q, thus expanding the set of priors q for which the pooling PBE can be supported.

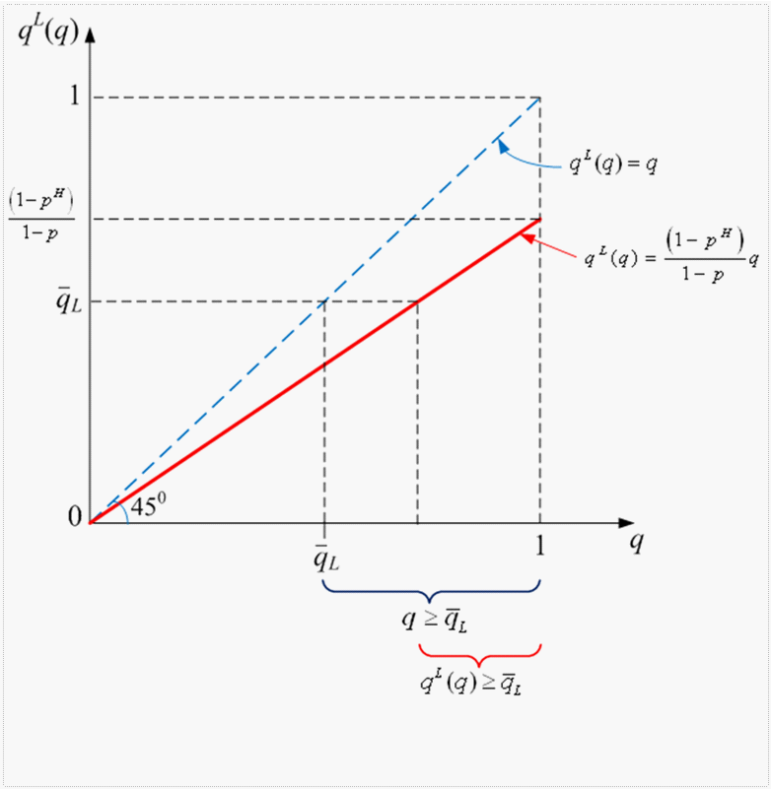

Figure A2

Separating PBE. Let us now analyze the separating PBE which, under uncorrelated types, can be sustained for all priors q satisfying $q < \overline{q}_L$. From our above discussion, note that $q < \overline{q}_L$ can be supported under a more restrictive set of priors q than condition $q^L(q) < \overline{q}_L$; as depicted in figure A2. Hence, the separating PBE can be sustained under a larger set of priors q when types are

positively correlated than otherwise. In contrast, when countries' types are negatively correlated, $p^H(p) < p < p^L(p)$ and $q^H(q) < q < q^L(q)$, inducing the low-type leader to participate under a larger set of priors q, thus shrinking the set of priors q for which the separating equilibrium can be sustained.

Semiseparating PBE. From our previous discussion, we can conclude that: (1) the set of priors q satisfying $q \geq \overline{q}_L$ under uncorrelated types is larger than that satisfying $q^L(q) \geq \overline{q}_L$, thereby restricting the set of priors q for which the semiseparating equilibrium can be supported when types are correlated; and (2) the set of priors p satisfying $p \geq \overline{p}_H$ under uncorrelated types is more restrictive than that satisfying $p^H(p) \geq \overline{p}_H$, thus expanding the set of priors p for which the semiseparating PBE can be sustained under correlated types. However, since the set of priors q under which the leader signs is more restrictive when types are positively correlated, the semiseparating PBE can be sustained under a more restrictive set of priors q. When countries' types are negatively correlated, $q^H(q) < q < q^L(q)$, inducing the low-type leader to participate under a larger set of priors q, thus expanding the set of priors q for which the semiseparating PBE can be sustained. ■

References

[1] Bagwell, K. (2009) "Self-Enforcing Trade Agreements and Private Information," NBER Working Paper, 14812.

[2] Barrett, S. (1994a) "Self-enforcing international environmental agreements," Oxford Economic Papers, 46, pp. 878–94.

[3] Barrett, S. (1994b) "Strategic Environmental Policy and International Trade," Journal of Public Economics, 54, pp. 325–38.

[4] Barrett, S. (1999) "A theory of full international cooperation," Journal of Theoretical Politics, 11, 519–41.

[5] Busch, M. L., and E. Reinhardt. (2003) "Developing Countries and GATT/WTO Dispute Settlement," Journal of World Trade, 37(4), pp. 719–35.

[6] Carraro, C., and D. Siniscalco. (2001) Transfers, Commitments and Issue Linkage in International Environmental Negotiations, in A. Ulph, ed., *Environmental Policy, International Agreements and International Trade*, Oxford University Press.

[7] Cesar, H. (1994) *Control and game models of the greenhouse effect*, Berlin: Springer-Verlag.

[8] Cho, I. and D. Kreps (1987) "Signaling games and stable equilibrium," Quarterly Journal of Economics, 102, 179-222.

[9] Downs, G.W., D.M. Rocke, and P.N. Barsoom (1996) "Is the good news about compliance good news about cooperation?" International Organization, 50, pp. 379–406.

[10] Ederington, J. (2002) "Trade and Domestic Policy Linkage in International Agreements," International Economic Review, 43(4), pp. 1347–68.

[11] Espinola-Arredondo, A. and F. Munoz-Garcia (2011) "Free-riding in International Environmental Agreements: A Signaling Approach to Non-Enforceable Treaties," Journal of Theoretical Politics, 23 (1), pp. 111–34.

[12] Fearon, J. D. (1998) "Bargaining, Enforcement, and International Cooperation," International Organization, 52 (2), pp. 269–305.

[13] Iida, K. (1993) "When and How Do Domestic Constraints Matter?: Two-Level Games with Uncertainty," Journal of Conflict Resolution, 37, pp. 403-26.

[14] Keohane, R. O. (1984). *After Hegemony: Cooperation and Discord in the World Political Economy.* Princeton, N.J.: Princeton University Press.

[15] Koremenos, B. (2001) "Loosening the ties that bind: A learning model of agreement flexibility," International Organization, 55(2), pp. 289–325.

[16] Koremenos, B., C. Lipson, and D. Snidal (eds.) (2004) *The Rational Design of International Institutions.* Cambridge: Cambridge University Press.

[17] Koremenos B. (2005). "Contracting around international uncertainty," American Political Science Review, 99, pp. 549–65.

[18] Lake, D. (1999) *Entangling Relations: American Foreign Policy in its Century.* Princeton, NJ: Princeton University Press.

[19] Lee, G. M. (2007). "Trade Agreements with Domestic Policies As Disguised Protection," Journal of International Economics, 71(1), 241-59.

[20] Martin, A. and W. Vergote (2008), "On the Role of Retaliation in Trade Agreements," Journal of International Economics, 76(1), 61-77.

[21] Martin, L. (2005) "The President and International Commitments: Treaties as Signaling Devices," Presidential Studies Quarterly, 35(3), pp. 440-65.

[22] Rosendorff, B.P. (2005) "Stability and rigidity: Politics and design of the WTO's dispute settlement procedure," American Political Science Review, 99(3), pp. 389–99.

[23] Schreurs, M. A., H. Selin, and S.D. VanDeveer (2009) "Expanding transatlantic relations: Implications for environment and energy policies," in M. A. Schreurs, H. Selin, & S. D. VanDeveer (Eds.), *Transatlantic environment and energy politics: Comparative and international perspectives* (pp. 1–18). Aldershot, UK: Ashgate Publishing.

[24] Stoett, P. J. (1997) The International Politics of Whaling, University of Washington Press.

[25] Tarar, A. (2001) "International Bargaining with Two-Sided Domestic Constraints," Journal of Conflict Resolution, 45(3), pp. 320-40.

[26] Von Stein, J. (2008). "The International Law and Politics of Climate Change: Ratification of the United Nations Framework Convention and the Kyoto Protocol," Journal of Conflict Resolution, 52(2), pp. 243-68.

[27] Whalley, J. (1991) "The Interface Between Environmental and Trade Policies," The Economic Journal, 101, pp. 180-89.

[28] Yarbrough, B.V. and R.M. Yarbrough (1992) *Cooperation and Governance in International Trade.* Princeton, NJ: Princeton University Press.

[29] Young, O. P. (1991) "Political Leadership and Regime Formation: On the Development of Institutions in International Society," International Organization, 45 (3), pp. 281–308.