Free-riding in International Environmental Agreements:

A Signaling Approach to Non-Enforceable 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

November 2, 2010

Abstract

This paper examines countries' free-riding incentives in international environmental agreements (IEAs) when, first, the treaty is non-enforceable, and second, countries do not have complete information about other countries' noncompliance cost. We analyze a signaling model whereby the country leading the negotiations of the international agreement can reveal its own noncompliance costs through the commitment level it signs in the IEA. Our results show that countries' probability to join the IEA is increasing in the free-riding benefits they can obtain from other countries' compliance, and decreasing in the cost of not complying with the initial

[*]We would like to thank participants of the Second Workshop on Game Theory in Energy, Resources and the Environment at HEC Montréal and the 17th Annual Conference of the EAERE in Amsterdam for their helpful comments and suggestions. We also thank the editor and two anonymous referees for their thoughtful comments.

[†]Address: 111C Hulbert Hall, Washington State University, Pullman, WA 99164. E-mail: anaespinola@wsu.edu. Phone: (509) 335 8494. Fax: (509) 335 1173.

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

terms of the agreement. This paper shows that, when free-riding incentives are strong enough, there is no equilibrium in which all countries join the IEA. Despite not joining the IEA, however, countries invest in clean technologies. Finally, we relate our results with some common observations in international negotiations.

KEYWORDS: Signaling games, environmental agreements, nonbinding negotiations, noncompliance cost.

JEL CLASSIFICATION: C72, D62, Q28.

1 Introduction

In recent years many industrialized nations have actively participated in numerous international environmental agreements (IEAs henceforth) to address the issues related to environmental degradation, from ozone layer depletion (in the Montreal protocol) to greenhouse gases (in the Kyoto protocol). By entering in these environmental treaties, countries accept a commitment level about how much they will invest in clean technologies, reduce pollutant emissions, etc. One important characteristic of IEAs is that commitment levels are non-enforceable, since there is no international organization which can perfectly enforce the content of the agreement. The non-binding nature of these treaties makes them particularly interesting from a theoretical perspective, since most of the negotiations analyzed by the literature on contract theory are binding and enforceable, such as in bargaining models. In addition, they are attractive from an applied approach, given the significant consequences that not implementing the agreements' content have on the environmental quality of the planet.

Because environmental treaties are non-enforceable, they cannot be treated as binding negotiations since countries can deviate from the terms of the agreement. A country would do so if complying with the full content of the agreement is too costly because of its own economic or political context. Intuitively, countries' political costs of not complying the treaty might be explained as the incumbent political party being severely punished by environmentally oriented citizens ("green voters"), whereas sticking to the terms of the agreement might be rewarded by these voters' support in future elections.[1] Therefore, countries' incentives to comply (or deviate from) the environmental agreement can be analyzed as a *strategic* decision. We henceforth refer to the political cost of not complying with the terms of the IEA as "noncompliance cost."

Furthermore, the signature of a specific commitment level in the IEA is not a simultaneous

negotiation, but rather, the result of a long and sequential negotiation process.[2] In addition, we assume that countries do not have complete information about other countries' noncompliance costs. Because of this sequential structure, it is common that countries base their decision on whether to join the IEA after observing if other countries joined the agreement as well. Particularly, a country's decision to join the agreement might reveal information about the country's own political cost from noncomplying the treaty in later stages, which ultimately affects other countries' decision on whether or not to sign the agreement.[3]

This paper examines the negotiation and implementation of IEAs as a signaling game where, first, the country leading the negotiations decides to sign (or not sign) the agreement. Afterwards, observing the leader's decision, the other country (follower) decides whether to join the treaty. Finally, both countries non-cooperatively and independently choose whether to comply with the terms of the agreement (e.g., investing in clean technologies or reducing pollutant emissions). That is, we analyze how the country's signature of an international agreement can reveal information about that country's private political situation to other countries, and how this information might affect other countries' decision to join the treaty.

We first identify the investment in clean technologies that, in the equilibrium of the second stage of the game, every country implements both when it joins the treaty and when it does not. We show that the investment increases in the commitment level that a country signs in the agreement, decreases in other countries' commitment level, and increases in the political cost that the country would suffer because of not fulfilling the treaty.

Additionally, we identify conditions under which the follower (the second mover in this signaling game) decides to join the IEA. We demonstrate that the follower decides to participate in the agreement if the environmental *benefits* arising from the IEA offset the noncompliance *cost* of

deviating from the agreement. Specifically, we show that a country's decision to join the agreement is more likely:

1. the higher the return from the improved environmental quality resulting from the treaty;
2. the lower the commitment level that the agreement specifies to the follower;
3. the higher the commitment level that the leading country signs in the IEA; and
4. the higher the probability that the leading country implements most of the commitments agreed to in the treaty.

These four incentives, especially (3) and (4), emphasize the fact that the second mover's decision to participate in the IEA can be rationalized as a free-riding behavior. That is, a country is more likely to join agreements in which other countries bear most of the burden of the investment in clean technologies.

We show that the unique equilibrium of the game (involving the use of pure strategies) is that in which the leading country signs the IEA when its noncompliance costs are low, but does not when its costs are high, and the following country responds by not signing the treaty. This constitutes a separating equilibrium in which the information about the country leading the negotiations is *fully transmitted* to the following country. We then strengthen our results by showing that this strategy profile can be supported as the *unique* equilibrium of the game for all parameter values and under relatively general assumptions on the utility function. In addition, we demonstrate that our results hold even when countries obtain a political benefit from the mere signature of the agreement, as long as such benefit does not offset the political cost from deviating from the treaty afterwards.

Finally, we provide an "optimistic" result: our equilibrium predictions show that certain countries with high political costs might decide not to join a particular IEA but nonetheless invest

positive amounts in clean technologies (or reduce emissions) afterwards. This result would explain the decision of countries like the United Kingdom and U.S.A. of not ratifying the Helsinki and Kyoto protocols respectively, but subsequently complying with many of the environmental goals included in these treaties.

The structure of the paper is as follows. First, we describe the literature on international agreements. Section two presents the model, and section three analyzes the second stage of the game, when countries decide how much to invest in clean technologies. Section four analyzes countries' incentives to join the agreement, and the unique equilibrium of the signaling game. Section five extends the findings to more general utility functions and parameter values. Section six concludes and provides a discussion about our results.

1.1 Related literature

In recent years, many authors have analyzed the negotiation stage of different IEAs using the theory of repeated games; see Barrett (1994a, 1994b, 1999), Cesar (1994), and Rubio and Ulph (2007). This literature considers that a country individually decides to join an international agreement if the benefits that it obtains from fully implementing the agreement outweigh the costs.[4] However, they assume that signatory countries comply the terms of the IEA. In contrast, this paper allows for the possibility of noncompliance, bringing our model and results closer to countries' observed behavior after signing certain international agreements in an incomplete information context.[5] Specifically, our findings predict the non-signature of the agreement by countries with high noncompliance costs. Hence, our results provide an explanation for countries' hesitation to participate in nonbinding IEAs such as the Kyoto protocol. Similar to the existing literature, our model identifies the existence of free-riding incentives in the negotiation stage of the agreement. However, this free-riding result is explained by countries' political cost of deviating from the content of the treaty.

Some studies have introduced incomplete information in environmental games. In particular, Bac (1996) develops a two country dynamic game where countries do not know each others' environmental valuations. He assumes that both countries simultaneously decide their abatement decisions, not allowing for the transmission of information in signaling games, and considering that countries' payoff structure is strategically equivalent to a "chicken game", unlike most literature on IEAs. Brandt (2004), in contrast, develops a signaling game where countries' incentives resemble those in a prisoner's dilemma game. Like these papers, we investigate countries' incentives to participate in IEAs under incomplete information. However, our model allows countries to deviate from their terms of the treaty.

Recent literature has approached the negotiation stage of international agreements –not only of environmental nature but also with political or economic content– introducing the assumption that countries are incompletely informed about each others' internal situation. For instance, Iida (1993) formalized Putman's (1988) model of international negotiations. In particular, Iida (1993) considers a repeated bargaining game in which countries do not have information about each others' domestic situation, and are allowed to make offers and counteroffers. Importantly, this paper assumes that once an offer has been agreed upon (signature of the treaty), both countries fully implement its content. As aforementioned, IEAs are rarely binding, which implies that the relative fulfillment of an agreement should be an equilibrium result rather than an assumption of the model. Morrow (1991) uses a similar approach as Iida (1993), but applies his model to the particular case of USA-Soviet Union negotiations under incomplete information.[6] This model focuses only on the benefits or costs that politicians can obtain from signing international agreements (as if the signature had a *per se* value), but is silent to the political benefits or costs from complying the terms of the agreement (or deviating from it). We introduce both types of incentives in our model. Finally, Bagwell (2009) develops a repeated game where countries sign a tariff agreement which can be

weakly binding, under incomplete information among countries. Countries suffer a political cost from setting the low tariffs specified in the treaty, unlike our model where countries benefit from complying the treaty.

2 Model

We represent the signature and implementation of an IEA as a two-stage game. For simplicity, we assume that the IEA is being negotiated by two countries (country i and j). The first stage of the game, the "negotiation stage," is a signaling game where country i announces its participation in the IEA, given a particular non-binding commitment towards the investment in emission-reducing technologies, c_i, determined by an international agency, such as the United Nations' Intergovernmental Panel on Climate Change. Conditional on this announcement, country j responds by determining its participation in the IEA, for a given commitment level[7], c_j. In the second stage of the game, referred to as the investment game, countries simultaneously choose their investment level in clean technology. In particular, the time structure of the game is the following:

1. Nature selects country i's political cost from not fulfilling the agreement, $\alpha_i \geq 0$, which is privately observed by country i but not by j. For simplicity, we assume that the political cost is either high ($\alpha_i = 1$) or low ($\alpha_i = 0$), with associated probabilities p and $1-p$, respectively.

2. After observing its own political cost, country i announces its participation in the IEA, for a given non-binding commitment level c_i.

3. After observing whether country i participates in the agreement, country j decides its signature of the treaty, for a given commitment level c_j, given its posterior beliefs about country i's political cost. We assume that country j's political cost is high, $\alpha_j = 1$, which is common knowledge among the players.[8] Furthermore, we assume that country j's non-signature of

the agreement implies that the IEA is not implemented. (We relax both assumptions in the last section of the paper, where we extend our results to all parameter values for both the leader and the follower, and to general utility functions).

4. If both countries participate in the IEA, they play a simultaneous-move game in which they determine the investment levels in emission-reducing technologies, x_i and x_j, that are finally implemented. If either country does not sign the treaty, commitment levels are $c_i = c_j = 0$, and countries select investment levels x_i and x_j accordingly.

Let $\mu(H|S)$ and $\mu(H|NS)$ denote country j's posterior beliefs about country i's high political cost of deviating from the treaty after observing that country i signed the IEA or that it did not, respectively. Moreover, assume that country i's utility function is represented by the following quasilinear utility function,

$$u_i(x_i, x_j, c_i) = \ w - x_i + \ln\left[m(x_i + x_j) + \alpha_i\left(x_i - c_i\right)\right] \tag{1}$$

In particular, the first term, $w - x_i$, represents the utility derived from the consumption of the remaining monetary units that are not invested in clean technologies, i.e., not invested in the public good. In the second term, m represents the return from the environmental good and $m(x_i + x_j)$ denotes the total return that country i obtains from the consumption of a higher level of environmental quality given its own investments, x_i, and the ones of country j, x_j. Finally, $\alpha_i\left(x_i - c_i\right)$ represents the return that country i derives from relatively fulfilling its commitment c_i in the environmental agreement or the cost that it incurs from noncomplying the agreement. Intuitively, an increase in country i's investment, x_i, has a traditional *public good* dimension, via $m(x_i + x_j)$, and an additional *fulfillment* dimension, via $\alpha_i\left(x_i - c_i\right)$.

3 Equilibrium investments

We first obtain both countries' equilibrium investments in the second stage of the game using backward induction, for any profile of commitment levels signed during the first stage of the game (which we discuss below), and for any profile of political costs (α_i, α_j), from Espinola-Arredondo (2009).

Lemma 1. *In the investment game, every country i's Nash equilibrium investment in emission-reducing technologies is*

$$x_i^* = \begin{cases} 1 + \frac{\alpha_i c_i}{m+\alpha_i} \ \textit{if} \ \alpha_i > \hat{\alpha}_i(\alpha_j) \\ \frac{\alpha_i(1+c_i)(\alpha_j+m)-\alpha_j m c_j}{\alpha_j m+\alpha_i(\alpha_j+m)} \ \textit{if} \ \alpha_i \in (\bar{\alpha}_i(\alpha_j), \hat{\alpha}_i(\alpha_j)] \\ 0 \ \textit{if} \ \alpha_i \in (0, \bar{\alpha}_i(\alpha_j)] \end{cases} \tag{2}$$

where $\hat{\alpha}_i(\alpha_j) = \frac{m c_j + \alpha_j(1+c_j)(m+c_j)}{(1+c_i)m}$ *and* $\bar{\alpha}_i(\alpha_j) = \frac{\alpha_j c_j m}{(1+c_i)(\alpha_j+m)}$.

In particular, country i's investment in clean technologies is at its maximum level when its concern about green voters, α_i, is sufficiently high, i.e., $\alpha_i > \hat{\alpha}_i(\alpha_j)$. When the importance that country i assigns to green voters decreases below $\hat{\alpha}_i(\alpha_j)$ its optimal investment also decreases, as the above lemma shows. That is, country i is not highly concerned about its own relative fulfillment of the IEA because it does not perceive the group of green voters as being relevant in future elections. Finally, if α_i drops below the threshold $\bar{\alpha}_i(\alpha_j)$, then its concerns about green voters' punishment are not strong enough to support any positive investment in clean technologies.[9]

4 Negotiation stage

Let us first examine country j's incentives to participate in the IEA after observing that country i did not sign the agreement.

Lemma 2. *If country i does not sign the IEA, country j will not sign the agreement, for any commitment level included in the IEA, and for any probability distribution over country i's noncompliance cost.*

In order to analyze the results in lemma 2, let us identify country j's costs and benefits of signing the IEA. On the one hand, country j's *costs* are political. That is, country j's noncompliance of the environmental agreement is punished by its voters in future elections. On the other hand, country j's *benefits* of participating in the IEA are related to the improvement in the global environmental quality due to its investment in clean technologies in the second stage of the game. However, in this case country i does not sign the IEA, which implies that the benefits arising from the treaty are mainly due to country j's own investments. In this situation the political costs of deviating from the terms of the agreement dominate the environmental benefits resulting from country j's investment. As a consequence, country j does not sign the IEA after observing that country i did not sign the agreement.

In contrast, when country i decides to participate in the IEA, country j's decision depends on the specific costs and benefits from the agreement, as the following lemma describes.

Lemma 3. *If country i signs the IEA, then country j also signs it if and only if*

1. $p > \frac{c_j}{\overline{c_j}(c_i,m)}$, *when all types of country i sign the IEA, or*
2. $\mu(H|S) > \frac{c_j}{\overline{c_j}(c_i,m)}$, *when all types of country i do not sign the IEA, or*
3. $c_j \leq \overline{c_j}(c_i,m)$, *when country i signs the IEA if and only if its political costs of reneging from the IEA are high, where* $\overline{c_j}(c_i,m) = (1+m)\left(\ln[1+2m+\frac{mc_i}{1+m}] - \ln[1+2m]\right)$.

Intuitively, country j's costs of participating in the environmental agreement are still purely political, as discussed above. However, the environmental benefits of signing the IEA (improved

environmental quality) are higher: country i signs the treaty for a given commitment level c_i, which *might* be implemented. Indeed, the environmental benefits of participating in the treaty are increasing in the probability that country i's political cost of deviating from the terms of the IEA is high, and in the commitment level that country i signs, c_i.

The environmental benefits that lead country j to participate in the IEA are also increasing in the return from an improvement in the environmental quality, m, reflecting the incentives to free-ride on country i's investment in clean technologies. In other words, country j's participation in the IEA can be supported for a larger set of parameter values, $\{c_i, c_j, p\}$, and beliefs $\mu(H|S)$ and $\mu(H|NS)$, the higher is the return from the global public good (improved world environmental quality). Finally, this set of parameter values inducing country j to sign the agreement becomes larger as c_j, the commitment level that the IEA specifies for country j, decreases. Intuitively, the lower is c_j the lower are country j's political costs of deviating from the agreement, which allows for more parameter combinations where country j's environmental benefits of signing the IEA are higher than its political costs. That is, a decrease in c_j strengthens country j's free-riding benefits from participating in the IEA.

Figures 1 and 2 depict the set of parameter values supporting country j's participation in the IEA, where the level set $c_j = \overline{c_j}(c_i, m)$ represents parameter values for which country j is indifferent between signing and not signing the treaty. Pairs of commitment levels (c_i, c_j) such that $c_j < \overline{c_j}(c_i, m)$, on the left hand side of the level set, illustrate IEAs that are signed by country j (its environmental benefits outweigh its political costs), while those commitment levels satisfying $c_j > \overline{c_j}(c_i, m)$, on the right hand side of the level set, represent IEAs in which country j would not participate.

Commitment levels supporting participation in IEAs

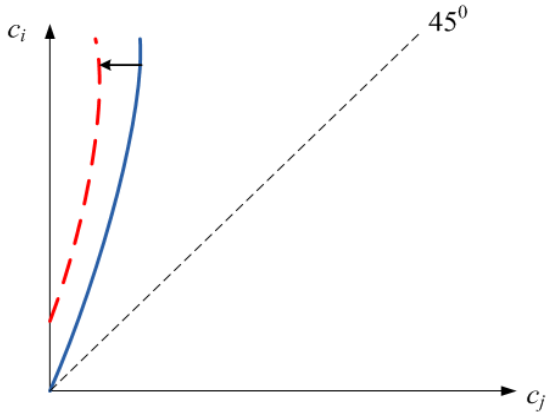


Figure 1. Level set $c_j = p \cdot \overline{c_j}(c_i, m)$ for $p = 0.9$ (solid), and $p = 0.7$ (dashed).

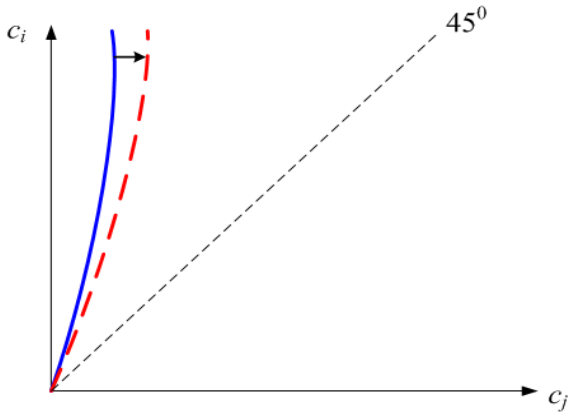


Figure 2. Level set $c_j = \overline{c_j}(c_i, m)$ for $m = 0.3$ (solid), and $m = 0.8$ (dashed).

In the case that country j cannot infer country i's type, the above level set is weighted by the probability that country i's commitment levels are likely to be implemented (the probability, p, that country i's noncompliance costs are high). The uncertainty about country i's type reduces country j's incentives to participate in the IEA. Figure 1 represents this effect by a leftward shift in level set $c_j = \overline{c_j}(c_i, m)$, which ultimately shrinks the set of commitment levels (c_i, c_j) for which country j accepts to participate in the environmental agreement. Finally, an increase in the return from the improved environmental quality, m, increases country j's incentives to free-ride on country i's investments during the second stage of the game, inducing the former to accept a larger variety of commitment levels. Figure 2 illustrates how increases in m provoke a rightward shift in the level set, enlarging the set of commitment levels for which country j decides to participate in the international agreement.

Other implication from the previous result is that the set of parameter values for which country

j accepts to participate in the IEA is above the 45^0-line where $c_i = c_j$. Intuitively, this implies that country j only signs an IEA if the commitment level suggested by the international agency, c_j, is lower than that recommended to country i, c_i, regardless of the return from the improvement in the global environmental quality resulting from the treaty.

4.1 Country i's equilibrium strategy

As shown in the previous sections, country j's incentives to participate in the IEA increase when its commitment level is low (since this reduces its political costs) and country i's commitment level is high (since this increases country j's environmental benefits from the improved environmental quality, i.e., free-riding). This result also applies to country i, the first mover in the signaling game, when country i is highly concerned about its political costs. Specifically, when it faces high political costs of deviating from the content of the IEA, country i would only be willing to participate in IEAs if it can benefit from large investments in clean technologies from country j (i.e., if free-riding incentives are strong enough), as the following lemma describes.

Lemma 4. *When country i's noncompliance costs are high, country i signs the IEA if and only if $c_i < \overline{c_i}(c_j, m)$. In contrast, if its noncompliance costs are low, country i participates in the IEA for any parameter values.*

On the one hand, when country i's noncompliance costs are high, country i signs if the environmental benefits arising from the implementation of the IEA during the second stage of the game offset the political costs. Importantly, country i's benefits from the IEA, as we mentioned before, are: (1) increasing in the other country's commitment level included in the IEA, c_j; (2) increasing in the return from the investment in clean technologies (return from the global public good, m); and (3) decreasing in the commitment level that the IEA specifies for country i. Intuitively, the above

three points can be jointly interpreted as country i's incentives to free-ride country j's investment in clean technologies.[10]

On the other hand, when country i faces a *low* political cost, it would be willing to participate in any type of IEA, i.e., for any parameter values c_i, c_j and m. Indeed, note that country i's benefits from signing the IEA are still arising from the improved environmental quality. This higher quality, nonetheless, is only due to country j's investment in clean technologies, since country i's optimal investment is zero during the second stage of the game.

The latter result has significant implications in our search for Perfect Bayesian Equilibria (PBE) of this signaling game. Indeed, it specifies that country i chooses to sign the IEA when its political costs are low, for any parameter values.[11] This implies that we only need to consider two possible PBE in pure strategies: the pooling PBE in which both types of country i decide to participate in the IEA; and the separating PBE in which the low concerned country signs, whereas the highly concerned country does not. We next show that one of these strategy profiles cannot be supported in equilibrium.

Lemma 5. *The pooling strategy profile in which both types of country i sign the IEA, regardless of their political costs, cannot be supported as a pure-strategy PBE of the environmental signaling game if* $c_j < p \cdot \overline{c_j}(c_i, m)$.

In a pooling strategy profile both types of country i would choose to sign the treaty. In this case, country j cannot update its beliefs about country i's type, and hence decides to join the IEA if and only if $c_j < p \cdot \overline{c_j}(c_i, m)$, where p denotes the prior probability that country i's political costs are high. However, this condition implies that country i would have to bear most of the burden of the treaty, leading it to not sign the agreement when its political costs are high, as shown in lemma

5.[12] This leaves the separating strategy profile as the *unique* PBE of this signaling game involving pure strategies, as the following proposition describes.

Proposition 1. *In the IEA signaling game, the following separating strategy profile can be supported as the unique PBE in pure strategies:*

1. *Country i signs the IEA when its political costs are low, but does not sign when its political costs are high, for any parameter values.*

2. *Country j responds by not signing the IEA, both after observing that country i signs and does not sign the agreement, for any parameter values, given that country j's posterior beliefs are $\mu(H|S) = 0$ and $\mu(H|NS) = 1$.*

Intuitively, this is a strategy profile in which country i's private information about its political costs is perfectly transmitted to country j. Particularly, when observing that country i signed the IEA, country j's beliefs assign full probability to country i having low political costs, whereas country i not signing the IEA reveals that country i's noncompliance costs are high.[13] Hence, country j knows that any positive commitment level comes from a country which will *not* be politically motivated to comply it. In turn, this eliminates country j's environmental benefit from participating in the IEA (i.e., the improved environmental quality that country j free-rides from country i's investment in clean technologies). Therefore, country j decides to not participate in the IEA for any parameter values.

Regarding the leader, the former proposition states that country i anticipates that country j will *not* sign the IEA, which reduces country i's environmental benefit from free-riding country j's investment in clean technologies. As a consequence, country i decides *not* to sign the IEA when its political costs are high. (Recall that, despite not participating in IEAs, countries might invest

positive amounts in clean technologies). In contrast, when its political costs are low, country i signs the IEA for any parameter values (lemma 4). Therefore, country i participates in the IEA if and only if its noncompliance costs are low.[14]

4.2 Discussion and applications

The results obtained in the previous section can illustrate usual negotiation patterns in IEAs. Indeed, countries with high political costs announce that they would only sign a high commitment level in the IEA (as the first movers in our signaling game) if other countries sign high commitment levels as well. However, as we just showed, countries with high political costs (both when they act as first and second movers) only accept to participate in the IEA if they can strongly benefit from other country's investment in clean technology during the second stage of the game (free-riding incentives). When both first and second mover countries bear high political costs, free-riding incentives are specially strong, inducing each country to only sign the IEA if the other country's commitment level is sufficiently high, relative to its own. Since this cannot simultaneously occur, both countries do not sign the IEA. Thus, the first mover's offer "I sign high commitment levels and you then sign high commitment levels as well", can be understood as a void proposal in equilibrium. The free-riding incentives can, however, be reduced by increasing countries' return from the environmental quality, m. This can be achieved through transfers of clean technology between countries –which reduce the investment cost for the benefiting country– or promoting policies that increase citizens' preference for renewable energies. In our model, both measures would expand the set of parameter values for which countries decide to participate in the agreement..

Our results show that, under incomplete information and political costs, countries do not participate in IEAs for any pair of commitment levels (c_i, c_j), including those that a central planner would select in order to maximize social welfare. Nonetheless, our model predicts that countries'

decentralized investment in clean technology during the second stage of the game is positive, suggesting that countries can still partially achieve their environmental goals despite not participating in IEAs.

Finally, our results can provide an interpretation about the relationship between the difficulty to monitor the compliance of certain IEAs and countries' observed behavior. Specifically, we could interpret countries' noncompliance costs in broader terms: including both the country's specific deviation from the environmental agreement *and* the probability that such deviation is detected (or perceived) by environmental agencies, NGOs and political parties running for office. Those IEAs that are particularly difficult to monitor are represented in our model by a reduction in the political costs that a signing country bears if it deviates from the terms of the agreement (lower α's). As described above, countries with low political costs will be willing to participate in IEAs, since the environmental benefits offset their (low) noncompliance costs.

5 Equilibria under general utility functions

In this section we extend the result obtained in proposition 1 to more general utility functions for country i and j. We show that the separating strategy profile in which country i participates in the IEA (when its noncompliance cost is low) is still the *unique* PBE of the signaling game for any parameter values, under relatively general conditions.

Let us denote by $U(c_i, c_j, \alpha_K)$ country i's equilibrium payoff from the second stage of the game (after including the equilibrium investment in clean technologies from both countries), where c_i and c_j denote, respectively, country i and j's commitment levels in the IEA, and α_K represents country i's noncompliance costs where $K = \{H, L\}$. When country i's noncompliance costs are high, we assume that $U(\cdot)$ is weakly decreasing in its own commitment level, c_i, for any given

commitment level of country j, i.e., $U(c_i, c_j, \alpha_H) \geq U(c_i', c_j, \alpha_H)$ for any $c_i < c_i'$. In contrast, when its noncompliance costs are low, $U(\cdot)$ is constant in c_i, for any given commitment level of country j, i.e., $U(c_i, c_j, \alpha_L) = U(c_i', c_j, \alpha_L)$ for any $c_i < c_i'$, since country i can deviate from the terms of the treaty without experiencing any political costs. Note that the above specification allows countries to obtain a benefit from the mere signature of the agreement. In particular, this benefit can arise in certain environmental agreements in which the country's participation produces a political gain because of an improved public image. Nonetheless, we consider that such benefit does not offset the political cost from deviating, and hence the above conditions still hold. Note that if, in contrast, the benefits from the mere signature of the IEA exceed the political cost from not complying the treaty, then countries would participate in the environmental agreement both when their political costs are low and high, supporting a pooling equilibrium in this game

Regarding country j, let $V(c_i, c_j, \alpha_K)$ represent its equilibrium payoff from the second stage of the game. Like in the case of country i, let us assume that $V(\cdot)$ is weakly decreasing in country j's commitment level c_j, for a given commitment level of country i, and for any noncompliance cost of country i, i.e., $V(c_i, c_j, \alpha_K) \geq V(c_i, c_j', \alpha_K)$ for any $c_j < c_j'$. Finally, we assume that both countries' equilibrium payoff weakly increases in the other country's commitment level. In the following proposition we show that, under the stated conditions, the unique PBE of the signaling game using pure strategies is that in which the leader participates in the IEA if and only if its noncompliance costs are low.

Proposition 2. *Assume the above conditions about the leader and the follower's second-period equilibrium payoffs. Then, in the IEA signaling game, the following (separating) strategy profile can be supported as the unique PBE in pure strategies:*

1. *Country i signs the IEA when its political costs are low, but does not sign when its political*

costs are high, for any parameter values.

2. *Country j responds by not signing the IEA both after observing that country i signs and does not sign the agreement, for any parameter values, given that country j's posterior beliefs are* $\mu(H|S) = 0$ *and* $\mu(H|NS) = 1$.

It is important to note the generality of the previous result. First, we assume that countries' equilibrium payoff weakly increases in each other's commitment levels (because of the environmental benefits that every country obtains free-riding other countries' investment in clean technologies), and weakly decreases in its commitment level (because of the noncompliance costs). As a consequence, we show that the separating equilibrium is the *unique* PBE in pure strategies: (1) for relatively general utility functions[15]; (2) for any noncompliance costs for country i and j (i.e., α_i and α_j do not need to take particular values); (3) without the need to grant veto power to any of the countries involved in the negotiation of the IEA; and (4) even if countries partially benefit from the mere participation in the treaty. Our results can then be used to analyze a variety of national and international agreements whose content cannot be perfectly enforced.

6 Conclusions

This paper examines countries' relative fulfillment of international environmental agreements (IEAs) when, first, the treaty is non-binding (lack of international organizations enforcing the terms of the agreement), and second, countries do not have complete information about how likely it is that other countries will stick to the terms of the agreement (or substantially deviate from it). We introduce the latter assumption by considering that countries experience a noncompliance cost, and that this cost is their own private information. We then construct a signaling model in which the country leading the negotiations of the international agreement can reveal its noncompliance cost

through its decision to accept certain commitment levels to be included in the IEA.

The paper first shows that countries' conflict of interest in this signaling game is especially strong. Indeed, every country decides to sign the agreement if the other country bears most of the burden of the treaty, if the return from the improved environmental quality is high enough, and if the probability that the other country will comply with its share of the treaty is high, or a combination of these three incentives. As we demonstrate, all these behavioral patterns can be rationalized from a free-riding perspective, since every country wants to benefit from the improved environmental quality arising from the agreement, but only if the investment is mostly carried out by *other* countries. We also show that such conditions guaranteeing the participation of every country in the treaty cannot be satisfied for both countries simultaneously (i.e., both countries will *not* join the IEA, under any parameter values).

As a result of the strong free-riding incentives, we show that the unique strategy profile that can be supported as a perfect Bayesian equilibrium of this game is a separating equilibrium, whereby information is perfectly transmitted from informed to uninformed countries. In particular, in this separating equilibrium the leading country in the negotiations signs the agreement only when its own political costs are low, and the following country does not sign the treaty for any parameter values. Note that our results are related to international agreements where western European countries leading the negotiations proposed a high reduction in their pollutant emissions, expecting other countries to join the agreement. As predicted by our equilibrium results, however, no other western country responded to such proposal by joining the IEA.

The paper raises two main implications. From a policy perspective, and despite the last (negative) result about countries' lack of participation in IEA, we show that many of the countries who decide to not join the treaty actually invest positive amounts in clean technologies, reduce their pol-

lutant emissions, etc. This is a positive result, and it is supported by the fact that several countries decide to invest in clean technologies despite their no participation in environmental treaties. Our result then helps to separate the commitment levels included in environmental agreements and their actual implementation, a common assumption in most of the literature on public and environmental economics. From a more theoretical approach, our results imply that when players (e.g., countries) negotiate non-binding contracts with incomplete information and with high competitive pressures (strong free-riding incentives), the unique strategy profile that can be supported as an equilibrium of the game is that in which players participate in the agreement only when its noncompliance costs are low. This is a powerful implication for other types of international agreements. Specifically, it suggests that in order to promote multiple cosignatories in IEAs, the most effective tool is to reduce countries' free-riding incentives arising during the subsequent implementation stage. More theoretical research is nonetheless needed in order to better understand the connection between non-binding international agreements and their latter implementation.

7 Appendix

7.1 *Semiseparating equilibria*

We know that when $c_j \geq p \cdot \overline{c_j}(c_i, m)$, country j does not participate in the IEA, regardless of country i's decision. Graphically, this condition is represented by the area to the right of the level set $c_j = p \cdot \overline{c_j}(c_i, m)$ in Figure 1. Indeed, parameter values in such region imply that either: (1) the commitment level for country j, c_j, is relatively high; or (2) the probability that country i will comply with the agreement, p, is low; or (3) the global environmental benefits from the treaty, m, are relatively low. Any combination of these three incentives induces country j to *not* participate in the treaty, even after observing that country i did. The leader (country i) is now nevertheless more willing to participate in the treaty, since the IEA specifies low commitment levels for country i (graphically represented by pairs to the right of the level set). As we summarize in the next result, the leader is now induced to randomize between signing and not signing the IEA, whereas the follower never participates, as suggested above.

Semiseparating equilibria. *In the IEA signaling game, the following strategy profile can be supported as a semiseparating PBE:*

1. *When country i's noncompliance costs are high, country i signs the treaty with probability*

$$q_H = \frac{c_j}{(m-1)p \cdot \widetilde{c_j}(c_i, m)} \tag{3}$$

where $\widetilde{c_i}(c_j, m) = \log(1+2m) - \log\left(1 + 2m + \frac{mc_i}{1+m}\right)$, *and when country i's noncompliance costs are low, country i signs the treaty with probability*

$$q_L = (1-p) + \frac{(1-p)c_j}{(1+m)\widetilde{c_j}(c_i, m)} \tag{4}$$

2. *Country j responds by not signing the IEA, regardless of country i's decision.*

Proof. Let us start with country j. After observing that country i signs, country j randomizes if and only if his beliefs $\mu(H|S) = \frac{pq_H}{pq_H+(1-p)q_L}$ satisfy

$$\begin{aligned} & \frac{pq_H}{pq_H+(1-p)\,q_L}V(c_i,c_j,\alpha_H)+\left(1-\frac{pq_H}{pq_H+(1-p)\,q_L}\right)V(c_i,c_j,\alpha_L) \\ = \quad & \frac{pq_H}{pq_H+(1-p)\,q_L}V(c_i,0,\alpha_H)+\left(1-\frac{pq_H}{pq_H+(1-p)\,q_L}\right)V(c_i,0,\alpha_L) \end{aligned}$$

where $V(c_i,c_j,\alpha_K)$ denotes country j's utility level from country i and j's commitment level, when country j's belief about country i's noncompliance cost is α_K where $\alpha_K = \{\alpha_H,\alpha_L\}$, and evaluated at the optimal investment level found in lemma 1. As usual, we assume that $\alpha_H = 1$ and $\alpha_L = 0$. Solving for q_H,

$$q_H\left(q_L\right) = \frac{(-1+p)\,c_j q_L}{p\left((1+m)\,\widetilde{c_j}(c_i,m)+c_j\right)}$$

where $\widetilde{c_j}(c_i,m) = \log(1+2m) - \log\left(1+2m+\frac{mc_i}{1+m}\right)$. After observing that country i did not sign, country j participates in the treaty if and only if his beliefs $\mu(H|NS) = \frac{p\left(1-q_H\right)}{p(1-q_H)+(1-p)(1-q_L)}$ satisfy

$$\begin{aligned} & \frac{p\,(1-q_H)}{p\,(1-q_H)+(1-p)\,(1-q_L)}V(0,c_j,\alpha_H)+\left(1-\frac{p\,(1-q_H)}{p\,(1-q_H)+(1-p)\,(1-q_L)}\right)V(0,c_j,\alpha_L) \\ = \quad & \frac{p\,(1-q_H)}{p\,(1-q_H)+(1-p)\,(1-q_L)}V(0,0,\alpha_H)+\left(1-\frac{p\,(1-q_H)}{p\,(1-q_H)+(1-p)\,(1-q_L)}\right)V(0,0,\alpha_L) \end{aligned}$$

that is, if

$$q_L\left(q_H\right) = \frac{(1+m)\,(-1+p)\left(\log\left[1+m\right]-\log\left[1+m+\frac{mc_j}{1+m}\right]\right)+c_j\,(-1+pq_H)}{(-1+p)\left((1+m)\left(\log\left[1+m\right]-\log\left[1+m+\frac{mc_j}{1+m}\right]\right)+c_j\right)}$$

Solving for q_H and q_L simultaneously,

$$q_H = \frac{c_j}{(m-1)p \cdot \widetilde{c_j}(c_i, m)} \quad \text{and} \quad q_L = (1-p) + \frac{(1-p)c_j}{(1+m)\widetilde{c_j}(c_i, m)}$$

On the other hand, country i randomizes between signing and not signing the treaty when its noncompliance costs are high if and only if

$$r \cdot U(c_i, c_j, \alpha_H) + (1-r)U(c_i, 0, \alpha_H) = s \cdot U(0, c_j, \alpha_H) + (1-s)U(0, 0, \alpha_H) \tag{5}$$

where r (s) refers to the probability that country j participates in the IEA after observing that country i signed (not signed, respectively) the treaty, and $U(c_i, c_j, \alpha_K)$ denotes country i's utility level from country i and j's commitment level, when country i's noncompliance cost is α_K where $\alpha_K = \{\alpha_H, \alpha_L\}$, and evaluated at the optimal investment level found in lemma 1. Solving for r,

$$r(s) = \frac{((1+m)\, s \cdot \widetilde{c_j}(c_i, m))}{(1+m)\, \widetilde{c_j}(c_i, m) + c_i}$$

And when country i's noncompliance costs are low, country i randomizes between signing and not signing if and only if

$$r \cdot U(c_i, c_j, \alpha_L) + (1-r)U(c_i, 0, \alpha_L) = s \cdot U(0, c_j, \alpha_L) + (1-s)U(0, 0, \alpha_L)$$

Solving for r and s simultaneously we obtain $r = 0$ and $s = 0$. Hence, country j does not sign after observing that country i signed ($r = 0$) or after observing that country i did not sign ($s = 0$). ■

7.2 *Proof of Lemma 1*

Both players simultaneously submit their investments in emission-reducing technologies. Fixing country j's investment, x_j, country i's utility maximization problem becomes

$$\max_{x_i} \quad w - x_i + \ln\left[m(x_i + x_j) + \alpha_i\left(x_i - c_i\right)\right]$$

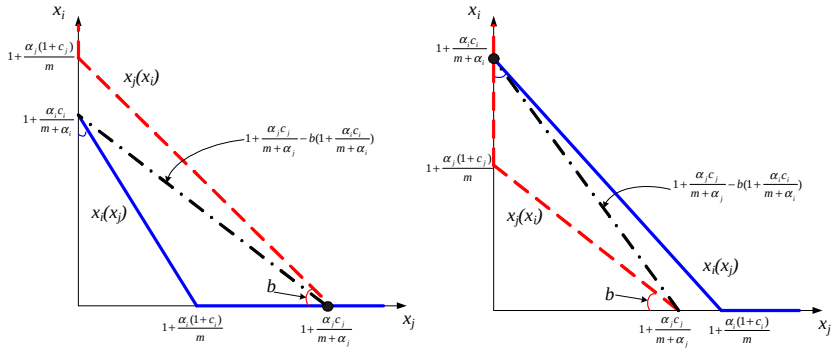
And the argument that maximizes this objective function gives us the following best response function

$$x_i(x_j) = \begin{cases} 1 + \frac{1}{m+\alpha_i}\alpha_i c_i \text{ if } x_j = 0 \\ 1 + \frac{1}{m+\alpha_i}\left[\alpha_i c_i - m x_j\right] \text{ if } x_j \in \left(0, \frac{\alpha_i(1+c_i)+m}{m}\right) \\ 0 \text{ if } x_j \geqslant \frac{\alpha_i(1+c_i)+m}{m} \end{cases}$$

Since $1 + \frac{1}{m+\alpha_i}\left[\alpha_i c_i - m x_j\right] = 0$ exactly at $x_j = \frac{\alpha_i(1+c_i)+m}{m}$. Hence, this best response function can be more compactly expressed as

$$x_i(x_j) = \begin{cases} 1 + \frac{1}{m+\alpha_i}\left[\alpha_i c_i - m x_j\right] \text{ if } x_j \in \left(0, \frac{\alpha_i(1+c_i)+m}{m}\right) \\ 0 \text{ if } x_j \geqslant \frac{\alpha_i(1+c_i)+m}{m} \end{cases}$$

Let us analyze the different forms in which country i and j's best response functions can cross each other. The corner solutions are illustrated in the following figures (cases 1 and 2 below), to clarify the following steps of the proof.



Case 1 Case 2

Case 1: $x_i^* = 0$

Note that $x_i^* = 0$ if and only the following two conditions are satisfied: (1) the horizontal intercept of country i's best response function is lower than that of country j, and (2) the slope of country j's best response function is small enough to make that $x_j(x_i)$ does not cross $x_i(x_j)$. That is, the first condition is satisfied if

$$\frac{\alpha_i(1+c_i)}{m} + 1 < \frac{\alpha_j c_j}{\alpha_j + m} + 1$$

Manipulating this inequality, we obtain

$$\alpha_i < \frac{\alpha_j c_j m}{(1+c_i)(\alpha_j + m)} \equiv \overline{\alpha_i}(\alpha_j) \tag{A.1}$$

On the other hand, the second condition holds if, b, the slope of country j's best response function, satisfies

$$0 < 1 + \frac{\alpha_j c_j}{m + \alpha_j} - b(1 + \frac{\alpha_i c_i}{m + \alpha_i})$$

$$\iff b < \frac{[m + \alpha_j(1+c_j)][m + \alpha_i]}{[m + \alpha_i(1+c_i)][m + \alpha_j]}$$

and since the slope of $x_j(x_i)$ is $\frac{m}{m+\alpha_j}$, we need that

$$\begin{aligned} \frac{m}{m+\alpha_j} &< \frac{[m+\alpha_j(1+c_j)][m+\alpha_i]}{[m+\alpha_i(1+c_i)][m+\alpha_j]} \\ [m+\alpha_i(1+c_i)][m+\alpha_j]m &< [m+\alpha_j(1+c_j)][m+\alpha_i][m+\alpha_j] \end{aligned}$$

and manipulating, and solving for α_i, we obtain the threshold of α_i below which all values of α_i support a zero investment in clean technologies by country i,

$$\alpha_i \leq \frac{mc_j + \alpha_j(1+c_j)(m+c_j)}{(1+c_i)m} \equiv \widehat{\alpha_i}(\alpha_j) \tag{A.2}$$

where $\overline{\alpha_i}(\alpha_j) < \widehat{\alpha_i}(\alpha_j)$, which implies that A.1 is more restrictive than A.2.

Case 2: $x_i^* = 1 + \frac{\alpha_i c_i}{m+\alpha_i}$

Let us now analyze the case in which country i sets the maximum investment $(1 + \frac{\alpha_i c_i}{m+\alpha_i})$, while country j does not invest. Firstly, we need that country i's horizontal intercept is above that of country j's, what implies

$$\frac{\alpha_i(1+c_i)}{m} + 1 > \frac{\alpha_j c_j}{\alpha_j + m} + 1 \Longleftrightarrow \alpha_i > \bar{\alpha}_i(\alpha_j)$$

Secondly, we need that b, the slope of country j's best response function, satisfies

$$0 > 1 + \frac{\alpha_j c_j + m}{m + \alpha_j} - b(1 + \frac{\alpha_i c_i}{m + \alpha_i})$$

and operating similarly as in the previous case, we have $\alpha_i > \hat{\alpha}_i(\alpha_j)$ And since $\overline{\alpha_i}(\alpha_j) < \widehat{\alpha_i}(\alpha_j)$, then $\alpha_i > \widehat{\alpha_i}(\alpha_j)$ is more restrictive.

Case 3: $x_i^* = \frac{\alpha_i(1+c_i)(\alpha_j+m) - \alpha_j m c_j}{\alpha_j m + \alpha_i(\alpha_j+m)}$

Finally, the equilibrium is interior when first, country i's horizontal intercept is below that of country j's, what implies,

$$\frac{\alpha_i(1+c_i)}{m}+1<\frac{\alpha_j c_j}{\alpha_j+m}+1 \Longleftrightarrow \alpha_i<\bar{\alpha}_i(\alpha_j)$$

and second, when b, the slope of country j's best response function, satisfies

$$0>1+\frac{\alpha_j c_j+m}{m+\alpha_j}-b(1+\frac{\alpha_i c_i}{m+\alpha_i}) \Longleftrightarrow \alpha_i>\hat{\alpha}_i(\alpha_j)$$

Hence, we can summarize the above three cases as follows:

$$x_i^*=\begin{cases} 1+\frac{\alpha_i c_i}{m+\alpha_i} \ \textit{if} \ \alpha_i>\hat{\alpha}_i(\alpha_j) \\ \frac{\alpha_i(1+c_i)(\alpha_j+m)-\alpha_j m c_j}{\alpha_j m+\alpha_i(\alpha_j+m)} \ \textit{if} \ \alpha_i\in(\bar{\alpha}_i(\alpha_j),\hat{\alpha}_i(\alpha_j)] \\ 0 \ \textit{if} \ \alpha_i\in(0,\bar{\alpha}_i(\alpha_j)] \end{cases}$$

where $\hat{\alpha}_i(\alpha_j)=\frac{mc_j+\alpha_j(1+c_j)(m+c_j)}{(1+c_i)m}$ and $\bar{\alpha}_i(\alpha_j)=\frac{\alpha_j c_j m}{(1+c_i)(\alpha_j+m)}$. ■

7.3 *Proof of Lemma 2*

If country j observes that country i signs a zero commitment level, then country j participates in the environmental agreement if and only if

$$\mu(H|NS)V(0,c_j,\alpha_H)+(1-\mu(H|NS))\,V(0,c_j,\alpha_L)\geq\mu(H|NS)V(0,0,\alpha_H)+(1-\mu(H|NS))\,V(0,0,\alpha_L)$$

where $V(c_i,c_j,\alpha_K)$ denotes country j's second-period equilibrium utility level from country i and j's commitment levels, when country i's type is $\alpha\in\{\alpha_H,\alpha_L\}$, and evaluated at the optimal investment level we found in lemma 1. This inequality holds only for $c_j\leq 0$. Given that $c_j>0$

by definition, then country j does not sign after observing no sign. Since this result does not depend on country j's beliefs about country i's political costs being high, $\mu(H|NS)$, then we can conclude that country j does not sign the agreement after observing that country i did not, for any probability distribution about country i's noncompliance costs. ■

7.4 *Proof of Lemma 3*

When both types of country i sign a positive commitment level, country j participates in the agreement after observing that country i signs (on the equilibrium path) if and only if

$$p \times V(c_i, c_j, \alpha_H) + (1-p) \times V(c_i, c_j, \alpha_L) \geq p \times V(c_i, 0, \alpha_H) + (1-p) \times V(c_i, 0, \alpha_L)$$

which holds if and only if $p \geq \frac{c_j}{\overline{c_j}(c_i,m)}$, where $\overline{c_j}\,(c_i,m) = (1+m)\left(\log\left[1+2m+\frac{mc_i}{1+m}\right] - \log\left[1+2m\right]\right)$. Similarly, when no type of country i participates in the international agreement, and country j observes the (off-the-equilibrium) action in which country i signs the treaty, then country j participates in the agreement if and only if

$$\mu(H|S)\times V(c_i, c_j, \alpha_H)+(1-\mu(H|S))\times V(c_i, c_j, \alpha_L) \geq \mu(H|S)\times V(c_i, 0, \alpha_H)+(1-\mu(H|S))\times V(c_i, 0, \alpha_L)$$

That is, if $\mu(H|S) \geq \frac{c_j}{\overline{c_j}(c_i,m)}$. When country j believes that country i signs the treaty only when its noncompliance costs are high, i.e., $\mu(H|S) = 1$ and $\mu(H|NS) = 0$, then country j signs after observing the signature of country i if and only if $V(c_i, c_j, \alpha_H) \geq V(c_i, 0, \alpha_H)$, that is, if $c_j \leq \overline{c_j}\,(c_i,m)$. Finally, if country j believes that country i signs the IEA only when its noncompliance costs are low, i.e., $\mu(H|S) = 0$ and $\mu(H|NS) = 1$, then country j signs after observing that country i signs the agreement if and only if $V(c_i, c_j, \alpha_L) \geq V(c_i, 0, \alpha_L)$, that is, if $c_j \leq 0$, i.e., country j does not participate in the treaty for any parameter values. ■

7.5 *Proof of Lemma 4*

High noncompliance costs. Let us consider the case in which the high type signs the treaty. In this case country i signs if and only if $U(c_i, c_j, \alpha_H) \geq U(0, 0, \alpha_H)$, where $U(\cdot)$ is country i's utility level from country i and j's commitment levels and evaluated at the optimal investment level we found in lemma 1. Note that $U(c_i, c_j, \alpha_H) \geq U(0, 0, \alpha_H)$ is satisfied when $c_i < \overline{c_i}\,(c_j, m)$.

Low noncompliance costs. In the strategy profile in which the low type decides to join the treaty, country i signs the agreement if and only if $U(c_i, c_j, \alpha_L) \geq U(0, 0, \alpha_L)$ which is true for all $c_j > 0$. ■

7.6 *Proof of Lemma 5*

First, note that in the strategy profile in which both types of country i participate in the agreement beliefs are $\mu(H|S) = p$, since posterior beliefs cannot be updated with equilibrium behavior, and $\mu\,(H|NS) \in [0, 1]$ for any off-the-equilibrium action of "not sign". Let us first analyze country j's equilibrium responses in this strategy profile, given the above system of beliefs. If country j observes that country i signs, then country j participates in the treaty if and only if

$$p \times V(c_i, c_j, \alpha_H) + (1-p) \times V(c_i, c_j, \alpha_L) \geq p \times V(c_i, 0, \alpha_H) + (1-p) \times V(c_i, 0, \alpha_L)$$

That is, if $p \geq \frac{c_j}{\overline{c_j}(c_i, m)}$, where $\overline{c_j}\,(c_i, m) = (1+m)\left(\log\left[1 + 2m + \frac{mc_i}{1+m}\right] - \log\left[1 + 2m\right]\right)$. If country j observes the (off-the-equilibrium) message in which country i does not sign, then country j responds by not joining the treaty either, as we showed in lemma 2. Let us now analyze country i when $p \geq \frac{c_j}{\overline{c_j}(c_i, m)}$ (and country j participates in the agreement after observing country i signing the treaty). If country i's noncompliance costs are high, it signs the treaty if and only if $U(c_i, c_j, \alpha_H) \geq U(0, 0, \alpha_H)$, that is, if $c_i \leq \overline{c_i}\,(c_j, m)$. In the case that country i's noncompliance costs are low, country i signs

the treaty if and only if $U(c_i, c_j, \alpha_L) \geq U(0, 0, \alpha_L)$ which is satisfied for any parameter values. Since condition $c_i \leq \overline{c_i}\,(c_j,m)$ and $p \geq \frac{c_j}{\overline{c_j}(c_i,m)}$ cannot be simultaneously satisfied, the pooling strategy profile in which both types of country i sign the treaty cannot be supported as a PBE when $p \geq \frac{c_j}{\overline{c_j}(c_i,m)}$. In the opposite case, when $p < \frac{c_j}{\overline{c_j}(c_i,m)}$, we have that country j does not participate in the international agreement after observing that country i signed the treaty. In this case country i signs the agreement when its own noncompliance costs are high if and only if $U(c_i, 0, \alpha_H) \geq U(0, 0, \alpha_H)$, which induces country i to be indifferent between signing and not signing the treaty when $p < \frac{c_j}{\overline{c_j}(c_i,m)}$. Similarly, if country i's noncompliance costs are low, it participates in the treaty if and only if $U(c_i, 0, \alpha_L) \geq U(0, 0, \alpha_L)$, which also implies that country i is indifferent between signing and not signing the agreement when $p < \frac{c_j}{\overline{c_j}(c_i,m)}$. Hence, when $p < \frac{c_j}{\overline{c_j}(c_i,m)}$ the pooling strategy profile in which both types of country i sign the treaty cannot be supported as a PBE of the game either. ■

7.7 *Proof of Proposition 1*

First, note that in the strategy profile in which country i signs the treaty only when its own noncompliance costs are low, and does not when its noncompliance costs are high. As a consequence, country j assigns full probability to country i being high after observing that it did not sign the agreement, $\mu(H|NS) = 1$, and full probability to country i being low when country i signs the treaty, $\mu(H|S) = 0$. Given this system of beliefs, let us now analyze country j's equilibrium responses. If country j observes that country i signs, then country j participates in the treaty if and only if

$$w - \left(1 + \frac{c_j}{m+1}\right) + \log\left[m\left(1 + \frac{c_j}{m+1} + 0\right) + \left(1 + \frac{c_j}{m+1} - c_j\right)\right] \geq$$

$$w-\left(1+\frac{0}{m+1}\right)+\log\left[m\left(1+\frac{0}{m+1}+0\right)+\left(1+\frac{0}{m+1}-0\right)\right]$$

which is not satisfied for any parameter values. Hence, country j does not sign the agreement after observing that country i participates in the treaty. If in contrast country j observes that country i does not sign the IEA, we know from Lemma 2 that country j does not participate in the international agreement, for any parameter values. Let us now focus on country i. When its noncompliance costs are high, it does not sign the treaty (as prescribed in this strategy profile) if and only if

$$w-\left(1+\frac{c_i}{m+1}\right)+\log\left[m\left(1+\frac{c_i}{m+1}+1+\frac{c_j}{m+1}\right)+\left(1+\frac{c_i}{m+1}-c_i\right)\right]\leq$$

$$w-\left(1+\frac{0}{m+1}\right)+\log\left[m\left(1+\frac{0}{m+1}+1+\frac{c_j}{m+1}\right)+\left(1+\frac{0}{m+1}-0\right)\right]$$

which holds for any parameter values. Similarly, when country i's noncompliance costs are low, country i signs the treaty if and only if

$$w-(0)+\log\left[m\left(0+1+\frac{c_j}{m+1}\right)+0\right]\geq w-(0)+\log\left[m\left(0+1+\frac{0}{m+1}\right)+0\right]$$

which is satisfied for any parameter values. ■

7.8 *Proof of Proposition 2*

1. *Separating where Leader$_H$ signs and Leader$_L$ does not sign.* First, note that $\mu(H|S)=1$ and $\mu(H|NS)=0$. As a consequence, after observing that the leader signs $c_i=\overline{c}_i>0$, country j does not participate in the IEA accepting $c_j=\overline{c}_j>0$ since $V(\overline{c}_i,\overline{c}_j,\alpha_H)\leq V(\overline{c}_i,0,\alpha_H)$, and after observing that the leader did not sign , $c_i=0$, country j does not sign the IEA since $V(0,\overline{c}_j,\alpha_L)\leq V(0,0,\alpha_L)$. Regarding the leader, when $\alpha=\alpha_H$ country i does not participate

in the IEA since $U(\overline{c}_i, 0, \alpha_H) < U(0, 0, \alpha_H)$, which implies that this strategy profile *cannot* be supported as a separating PBE of the game.

2. *Separating where Leader$_H$ does not sign and Leader$_L$ signs.* First, note that $\mu(H|S) = 0$ and $\mu(H|NS) = 1$. As a consequence, after observing that the leader signs $c_i = \overline{c}_i > 0$, country j does not participate in the IEA since $V(\overline{c}_i, \overline{c}_j, \alpha_L) \leq V(\overline{c}_i, 0, \alpha_L)$, and after observing that the leader did not sign, country j does not participate in the agreement since $V(0, \overline{c}_j, \alpha_H) \leq V(0, 0, \alpha_H)$. Regarding the leader, when $\alpha = \alpha_H$ country i does not participate in the IEA since $U(\overline{c}_i, 0, \alpha_H) \leq U(0, 0, \alpha_H)$. In contrast, when $\alpha = \alpha_L$ country i signs since $U(\overline{c}_i, 0, \alpha_L) \geq U(0, 0, \alpha_L)$. Hence, this strategy profile *can* be supported as a separating PBE of the game.

3. *Pooling where both types of leading countries sign.* First, note that $\mu(H|S) = p$ and $\mu(H|NS) \in [0, 1]$. As a consequence, after observing that the leader signs, country j does not sign the IEA since

$$p \times V(\overline{c}_i, \overline{c}_j, \alpha_H) + (1-p) \times V(\overline{c}_i, \overline{c}_j, \alpha_L) \leq p \times V(\overline{c}_i, 0, \alpha_H) + (1-p) \times V(\overline{c}_i, 0, \alpha_L)$$

$$\Longleftrightarrow p\left[V(\overline{c}_i, \overline{c}_j, \alpha_H) - V(\overline{c}_i, 0, \alpha_H)\right] \leq (1-p)\left[V(\overline{c}_i, 0, \alpha_L) - V(\overline{c}_i, \overline{c}_j, \alpha_L)\right]$$

since $V(\overline{c}_i, \overline{c}_j, \alpha_H) \leq V(\overline{c}_i, 0, \alpha_H)$ and $V(\overline{c}_i, 0, \alpha_L) \geq V(\overline{c}_i, \overline{c}_j, \alpha_L)$. And after observing that the leading country i did not sign, country j does not participate in the IEA given that

$$\begin{aligned} & \mu(H|NS) \times V(0, \overline{c}_j, \alpha_H) + (1 - \mu(H|NS)) \times V(0, \overline{c}_j, \alpha_L) \\ \leq \;& \mu(H|NS) \times V(0, 0, \alpha_H) + (1 - \mu(H|NS)) \times V(0, 0, \alpha_L) \end{aligned}$$

which holds because of $V(0, \overline{c}_j, \alpha_H) \leq V(0, 0, \alpha_H)$ and $V(0, 0, \alpha_L) \geq V(0, \overline{c}_j, \alpha_L)$. Therefore,

country j responds by not participating in the IEA after observing any previous decision by country i. Regarding the leader, when $\alpha = \alpha_H$ country i does not sign since $U(\overline{c}_i, 0, \alpha_H) \leq U(0, 0, \alpha_H)$. Hence, the strategy profile in which all types of country i sign the agreement *cannot* be supported as a pooling PBE of the game.

4. *Pooling where both types of leading countries do not sign the IEA.* First, note that $\mu(H|NS) = p$ and $\mu(H|S) \in [0, 1]$. As a consequence, after observing that the leader signs the treaty, country j does not sign the IEA since

$$\begin{aligned} & \mu(H|S) \times V(\overline{c}_i, \overline{c}_j, \alpha_H) + (1 - \mu(H|S)) \times V(\overline{c}_i, \overline{c}_j, \alpha_L) \\ \leq \quad & \mu(H|S) \times V(\overline{c}_i, 0, \alpha_H) + (1 - \mu(H|S)) \times V(\overline{c}_i, 0, \alpha_L) \end{aligned}$$

which is satisfied because of $V(\overline{c}_i, \overline{c}_j, \alpha_H) \leq V(\overline{c}_i, 0, \alpha_H)$ and $V(\overline{c}_i, 0, \alpha_L) \geq V(\overline{c}_i, \overline{c}_j, \alpha_L)$. Similarly, after observing that the leading country i did not sign, country j does not participate in the IEA given that

$$p \times V(0, \overline{c}_j, \alpha_H) + (1 - p) \times V(0, \overline{c}_j, \alpha_L) \leq p \times \nu(0, 0, \alpha_H) + (1 - p) \times V(0, 0, \alpha_L)$$

$$p \left[V(0, \overline{c}_j, \alpha_H) - V(0, 0, \alpha_H) \right] \leq (1 - p) \left[V(0, 0, \alpha_L) - V(0, \overline{c}_j, \alpha_L) \right]$$

since $V(0, \overline{c}_j, \alpha_H) \leq V(0, 0, \alpha_H)$ and $V(0, 0, \alpha_L) \geq V(0, \overline{c}_j, \alpha_L)$. Then, the following country responds by not participating in the IEA after observing any previous decision from country i. Regarding the leader, when $\alpha = \alpha_H$ country i does not sign since $U(\overline{c}_i, 0, \alpha_H) \leq U(0, 0, \alpha_H)$, and when $\alpha = \alpha_L$ country i is indifferent between signing and not signing the treaty (country i randomizes) given that $U(\overline{c}_i, 0, \alpha_L) = U(0, 0, \alpha_L)$. Hence, we can conclude that the strategy profile in which both types of country i do not sign the agreement *cannot* be supported as a

pooling PBE in pure strategies. ■

References

[1] Bac, Mehmet (1996) "Incomplete Information and Incentives to Free Ride on International Environmental Resources," Journal of Environmental Economics and Managements 30: 301-15.

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

[3] Barrett, Scott (1992) "Convention on Climate Change—Economic Aspects of Negotiations," Paris, OECD.

[4] Barrett, Scott (1994a) "Self-Enforcing International Environmental Agreements," Oxford Economic Papers 46: 878-94.

[5] Barrett, Scott (1994b) "The Biodiversity Supergame," Environmental and Resource Economics 4: 111-22.

[6] Barrett, Scott (1999) "A Theory of Full International Cooperation," Journal of Theoretical Politics 11: 519-41.

[7] Brandt, Urs S. (2004) "Unilateral Actions, the Case of International Environmental Problems," Resource and Energy Economics 26: 373-91.

[8] Carraro, Carlo, and Domenico 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.

[9] Cesar, Herman S.J. (1994) Control and Game Models of the Greenhouse Effect. Berlin: Springer-Verlag.

[10] Chayes, Abram, and Antonia Chayes (1995) The New Sovereignty: Compliance with International Regulatory Agreements. Cambridge, MA: Harvard University Press.

[11] Espinola-Arredondo, Ana (2009) "Free-Riding and Cooperation in Environmental Games," Journal of Public Economic Theory 11 (1): 119–58. .

[12] Folmer, Henk, Pierre van Mouche and Shannon Ragland (1993) "Interconnected Games and International Environmental Problems," Environmental and Resource Economics 3: 313-35.

[13] Fudenberg, Drew, and Jean Tirole (1984) "The Fat-Cat Effect, the Puppy-Dog Ploy, and the Lean and Hungry Look," American Economic Review 74(2): 361–66.

[14] Gilligan, Michael J. (2004) "Is There a Broader-Deeper Trade-off in International Multilateral Agreements?," International Organization 58: 459-84

[15] Hoel, Michael and Kerstin Schneider (1997) "Incentives to Participate in an International Environmental Agreement," Environmental and Resource Economics 9: 153–70.

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

[17] Lange, Andreas and Carsten Vogt (2003) "Cooperation in International Environmental Negotiations Due to a Preference for Equity," Journal of Public Economics 87: 2049-67.

[18] Mo, Jongryn (1994) "The Logic of Two-Level Games with Endogenous Domestic Coalition," Journal of Conflict Resolution 38: 402-22.

[19] Morrow, James D. (1991) "Electoral and Congressional Incentives and Arms Control," Journal of Conflict Resolution 35: 245-65.

[20] Putnam, Robert D. (1988) "Diplomacy and Domestic Politics: The Logic of Two-Level Games," International Organization 42: 427-60.

[21] Rubio, Santiago and Alistair Ulph (2007) "An Infinite-horizon Model of Dynamic Membership of International Environmental Agreements," Journal of Environmental Economics and Management 54 (3): 296-310.

[22] Whalley, John (1991) "The Interface Between Environmental and Trade Policies," The Economic Journal 101: 180-89.

Notes

[1]For instance, the defeat of Australian Prime Minister John Howard in 2007 has been explained, among other factors, by his noncompliance of climate change policies, i.e., Australia signed the Kyoto protocol in April 1998, but neither ratified it nor implemented its content. Alternatively, political costs can be understood as the deterioration in the international image from the lack of compliance of the IEA. From an ethical perspective, the political cost can be interpreted as the disappointment of the country's voters with the politicians who were supposed to implement the content of the treaty, since they lied both to their constituents and to the countries participating in the international agreement.

[2]The Kyoto protocol, for instance, was open for signature March 16, 1998, but did not enter into force until February 16, 2005.

[3]We assume that countries only have information about the probability distribution with which other country's noncompliance costs are high or low, but do not observe the actual realization of this random variable. In our setting, this is equivalent to considering that countries know some information about other countries' political scenario (e.g., from international news agencies), but are unaware of the *specifics* of the actual political situation in that country.

[4]This literature was complemented by other class of models in which countries were assumed to exhibit a preference for "international equality", Lange and Vogt (2003) and Hoel and Schneider (1997), by introducing the possibility that international negotiations impose sanctions on "defecting" countries, Barret (1992, 1994a), or to link the negotiations of transboundary pollution with other issues such as free-trade agreements, Carraro and Siniscalco (2001), Whalley (1991) and Folmer et al (1993).

[5]Chayes and Chayes (1995) also recognize that noncompliance exists, but they explain it by relying on uncontrollable social or economic changes affecting the signatory country. Our paper adds to this explanation for noncompliance the fact that countries can deviate from the terms of the agreement as long as it is politically beneficial for them.

[6]He introduces incomplete information in the signing stage, since the USA does not know the reservation value of the Soviet Union if the latter were to break the negotiations. Additionally, he assumes that the incumbent politician in the USA can improve his probabilities of being reelected if he signs a treaty with the Soviet Union.

[7]Similar to Gilligan (2004) countries can also select different commitments levels in our model, $c_i \neq c_j$.

[8]Common knowledge about α_j can be rationalized when the follower is a country with a long history negotiating IEAs, while the leader is a country with a relatively short history as an IEA negotiator.

[9]Note that when voters do not care about deviations from the commitment levels included in the IEA, $\alpha_i = \alpha_j = 0$,

countries' participation in the IEA would act as cheap talk.

[10] Note that countries' investment levels are strategic substitutes, i.e., an increase in x_i reduces country j's incentives to raise x_j, for any $i \neq j$. Furthermore, recall that countries prefer to participate in IEAs that specify low commitment levels for themselves but high for other countries. This strategic setting would correspond to the "Lean and Hungry Look" case in Fudenberg and Tirole's (1984) taxonomy about the effect of strategic pre-commitment on firms' competition.

[11] This result is applicable to the case of strategy profiles supported as PBE if we restrict our attention to degenerated (pure) strategies. However, when we allow for countries to randomize between signing and not signing (in the semiseparating equilibrium of the game), we show that country i may decide not to sign the IEA, regardless of its political costs (see Appendix).

[12] In contrast, when $c_j > p \cdot \overline{c_j}(c_i, m)$ holds country j does not participate in the IEA, regardless of country i's previous decision, which leads country i to be indifferent between signing and not signing the agreement when its own political costs are high. We analyze this mixed strategy profile as one special case of all mixed strategy profiles that can be supported as PBEs of this signaling game (see Appendix).

[13] Note that such separating equilibrium, in which information is fully transmitted, coincides with the subgame perfect equilibrium where countries are informed about each other's political costs.

[14] The introduction of more potential participants in the IEA would not substantially modify our results as long as information from the informed country (leader) to the uninformed country (follower) is perfectly transmitted in the first stage of the game, as shown in Proposition 1.

[15] Note that we do not impose linearity or separability on either country's utility function.