Fondazione Eni Enrico Mattei

# NOTA DI LAVORO

2.2016

## Sharing R&D Investments in Breakthrough Technologies to Control Climate Change

**Santiago J. Rubio**, Department of Economic Analysis and ERI-CES, University of Valencia

Mitigation, Innovation and Transformation Pathways
Series Editor: Massimo Tavoni

# Sharing R&D Investments in Breakthrough Technologies to Control Climate Change

By Santiago J. Rubio, Department of Economic Analysis and ERI-CES, University of Valencia

## Summary

This paper examines international cooperation on technological development as an alternative to international cooperation on GHG emission reductions. In order to analyze the scope of cooperation, a three-stage technology agreement formation game is solved. First, countries decide whether or not to sign up to the agreement. Then, in the second stage, the signatories (playing together) and the non-signatories (playing individually) select their investment in R&D. In this stage, it is assumed that the signatories not only coordinate their levels of R&D investment but also pool their R&D efforts to fully internalize the spillovers of their investment in innovation. Finally, in the third stage, each country decides non-cooperatively upon its level of energy production. Emissions depend on the decisions made regarding investment and production. If a country decides to develop a breakthrough technology in the second stage, its emissions will be zero in the third stage. For linear environmental damages and quadratic investment costs, the grand coalition is stable if marginal damages are large enough to justify the development of a breakthrough technology that eliminates emissions completely, and if technology spillovers are not very important.

**Keywords:** International Environmental Agreements, R&D Investment, Technology Spillovers, Breakthrough Technologies

**JEL Classification:** D74, F53, H41, Q54, Q55

*The author would like to thank Paco André, Hassan Benchekroun and participants at the 2015 Conference on Global Harmonized Carbon Pricing: Looking beyond Paris (Yale Center for the Study of Globalization) and the 2015 AERNA Workshop on Game Theory and the Environment (Autonomous University of Madrid) for their stimulating discussion. The usual disclaimer applies. Financial support from the Spanish Ministry of Economics and Competitiveness under project ECO2013-45045-R, and Generalitat Valenciana under project PROMETEO II/2014/054 is gratefully acknowledged.*

*Address for correspondence:*
Santiago J. Rubio
Department of Economic Analysis
University of Valencia
Edificio Departamental Oriental
Avda. de los Naranjos s/n
46022 Valencia
Spain
E-mail: santiago.rubio@uv.es

The opinions expressed in this paper do not necessarily reflect the position of
Fondazione Eni Enrico Mattei
Corso Magenta, 63, 20123 Milano (I), web site: www.feem.it, e-mail: working.papers@feem.it

# Sharing R&D Investments in Breakthrough Technologies to Control Climate Change*

Santiago J. Rubio

Department of Economic Analysis and ERI-CES

University of Valencia, Spain

September 25, 2015

*The author would like to thank Paco André, Hassan Benchekroun and participants at the 2015 Conference on Global Harmonized Carbon Pricing: Looking beyond Paris (Yale Center for the Study of Globalization) and the 2015 AERNA Workshop on Game Theory and the Environment (Autonomous University of Madrid) for their stimulating discussion. The usual disclaimer applies. Financial support from the Spanish Ministry of Economics and Competitiveness under project ECO2013-45045-R, and *Generalitat Valenciana* under project PROMETEO II/2014/054 is gratefully acknowledged.

## Abstract

This paper examines international cooperation on technological development as an alternative to international cooperation on GHG emission reductions. In order to analyze the scope of cooperation, a three-stage technology agreement formation game is solved. First, countries decide whether or not to sign up to the agreement. Then, in the second stage, the signatories (playing together) and the non-signatories (playing individually) select their investment in R&D. In this stage, it is assumed that the signatories not only coordinate their levels of R&D investment but also pool their R&D efforts to fully internalize the spillovers of their investment in innovation. Finally, in the third stage, each country decides non-cooperatively upon its level of energy production. Emissions depend on the decisions made regarding investment and production. If a country decides to develop a breakthrough technology in the second stage, its emissions will be zero in the third stage. For linear environmental damages and quadratic investment costs, the grand coalition is stable if marginal damages are large enough to justify the development of a breakthrough technology that eliminates emissions completely, and if technology spillovers are not very important.



**Keywords:** International environmental agreements; R&D investment; Technology spillovers; Breakthrough technologies

**JEL Classification System:** D74, F53, H41, Q54, Q55

# 1 Introduction

A lot of research has been developed in recent years looking for a new architecture to build an effective international agreement to control climate change to replace the Kyoto Protocol. One of the lines of research has focused on international cooperation to develop cleaner technologies as a complement or indeed an alternative to cooperation in emissions abatement. The idea explored by this literature is if reductions in abatement costs as a result of R&D investments might increase the willingness of countries to undertake significant emissions abatement and participate in an emissions agreement.[1] Among the different contributions in this line, we would like to highlight an early paper by Carraro and Siniscalco (1997). They present a model in which the signatories' technology is a *club good*. In their paper the signatories of the agreement do not coordinate their levels of R&D investment but they commit to share R&D efforts and avoid duplication of R&D activities. As a result there is a higher degree of technology spillover within rather than outside the agreement.[2] The authors investigate the stability of an abatement agreement complemented by this type of technological cooperation and provide a numerical example for which the grand coalition is stable. However,they do not evaluate the gains obtained from such cooperation.

Later on, Barrett (2006) has made more precise this optimistic result on participation. He analyzes whether a system of two agreements, one promoting cooperative R&D investment in a breakthrough technology that is a *pure public good* and the other stimulating the collective adoption of the technology, can outperform an abatement agreement such as the Kyoto Protocol.[3] Assuming linear benefit and cost of adopting the new technology, his analysis yields a standard result in the literature on international environmental

[1] An overview of technology-oriented agreements stressing their potential role in addressing the free-riding incentives in climate negotiations can be found in de Coninck et al. (2008).

[2] In fact, they assume that the degree of spillover for the non-signatories is zero. Under this assumption the signatories' technology is a club good. A club good is a public good that becomes excludable. See Cornes and Sandler (1996) for an excellent presentation of the theory of club goods.

[3] A breakthrough technology opens the possibility of GHG emissions being completely eliminated, i.e. fossil fuels could be completely replaced by other non-polluting energies. See Barrett (2009) for a survey of the possibilities of developing this type of technology.

agreements (henceforth, IEAs): membership can be larger but only when the agreement does not make all countries substantially better off. Observe that this result does not contradict Carraro and Siniscalco's (1997) conclusion but qualifies it: full cooperation is possible but it does not necessarily come along with large welfare gains.[4]

In this paper, we present a model built on the work by Carraro and Siniscalco's (1997) to address the issue analyzed by Barrett (2006): the cooperation in the adoption of a breakthrough technology, and we show that full cooperation might indeed imply large welfare gains, thus overcoming the small coalition paradox. Although our work uses some of the ingredients in Carraro and Siniscalco's (1997) model, there are also some differences we would like to clarify. Firstly, the technology is not a club good but an *impure public good:* technology spillovers are larger between signatories but they are not zero for non-signatories. The signatories form an international research joint venture to fully internalize the spillover effects but we think that some technological diffusion among the non-signatories is inevitable. Secondly, countries only negotiate a technology agreement i.e. we focus on cooperation in the development of cleaner technologies as an *alternative* to cooperation in abatement, whereas Carraro and Siniscalco (1997) focus on cooperation in the development of cleaner technologies as a *complement* to cooperation in reduction of gas emissions. Thus, in our model, there will not be cooperation in abatement even if the signatories do not invest enough to develop a breakthrough technology. Finally, the signatories not only promote an international research joint venture but also coordinate their levels of R&D investment in order to maximize the aggregate net benefits of the agreement. To analyze the stability of the agreement we solve a coalition formation game in three stages. Firstly, countries decide whether or not to participate in the agreement. Then, in the second stage, signatories choose their R&D investment to maximize the aggregate net benefit of the agreement and pool their R&D efforts to fully internalize the spillover effects, taking as given the non-signatories' R&D investment. On the other hand, the non-signatories choose their R&D investment to maximize their national net

[4]Recently, Nordhaus in his Presidential Address at the 2015 meeting of the American Economic Association has called this result: the "small coalition paradox". See Nordhaus (2015). The paradox goes back to the seminal papers on IEAs by Carraro and Siniscalco (1993) and Barrett (1994).

benefit, also taking as given the R&D investment of all other countries. Finally, in the third stage, each country decides non-cooperatively on its level of energy production. Emissions depend on the decisions taken regarding the levels of investment and production. If the country decides to develop a breakthrough technology in the second stage, then emissions will be zero in the third stage.

Our findings show that when all countries participate in the agreement, R&D investment increases with marginal damages and that there exists a threshold value of marginal damages for which a zero-emissions technology is adopted. We also show that there exists a second threshold value for marginal damages that is larger than the former threshold value, such that for whatever marginal damages greater than this second threshold value, a technology agreement with full participation is stable provided that technology spillovers are not very great. Thus, cooperation is successful when it is more needed i.e. when marginal damages are large. This means that the grand coalition may be stable but only when marginal damages are large enough for the development of a breakthrough technology makes economic sense. Stability comes along with the development of a breakthrough technology. We find that for marginal damages larger than this threshold value, when one country leaves the grand coalition it selects a level of R&D investment greater than the level it selected acting as a signatory, supporting then larger investment costs. One the one hand, if the country decides to continue using a breakthrough technology once it has left the agreement, it will have to invest more because of the drop in the spillover effects from signatories. Not being part of the technology agreement has a cost in terms of lower spillovers. On the other hand, if the country decides to develop a polluting technology, it will also have to invest more because in that case the optimal investment of the non-signatories depends on marginal damages, whereas this is not the case for the signatories. Marginal damages are critical to decide whether the signatories will develop a breakthrough technology or not. However, they do not affect the investment costs of developing this type of technology. What our analysis shows is that there is a threshold value for marginal damages above which the investment of the country that leaves the agreement is larger than the investment it supported acting as a signatory of a technology agreement with full participation, even if this country develops a polluting technology. Then, the grand coalition is stable since marginal damages cannot decrease with the exit of the country, and the energy production of the country leaving the agreement cannot increase. Thus, if one country leaves the grand coalition and the rest of signatories stay in the agreement, its net benefit decreases and the grand coalition is internally stable. Finally, we would like to point out that we derive a sufficient condition for the stability of the grand coalition. However, we expect the grand coalition to be stable for lower values of marginal damages that do not satisfy this sufficient condition.

Although the literature on IEAs is extensive, only papers by Hoel and de Zeeuw (2010, 2014) and El-Sayed and Rubio (2014) have focused on the stability analysis of a technology agreement. The rest of the literature includes papers by Ruis and de Zeeuw (2010), Hong and Karp (2012), Urpelainen (2014), Battaglini and Harstad (2015), Goeschl and Perino (2015) and Helm and Schmidt (2015). These authors consider countries negotiating an emissions agreement and study the impact that investment in cleaner technologies can have on the participation and effectiveness of the agreement including in the analysis the possibility that countries also cooperate in the development of a clean technology. Hoel and de Zeeuw (2010, 2014) develop a linear model where abatement costs decrease with the amount of total R&D investments by all countries. The new technology is hence considered to be a perfect public good. Moreover, they assume that there is no cooperation when countries in the last stage decide on abatement. They show in their 2014 model with heterogeneous countries that there may be an equilibrium with signatories investing in R&D and with non-signatories adopting the new technology. This means that a focus on technology development in IEAs may be successful in terms of abatement without the need for a large membership. However, they find that the relationship between aggregate abatement and the countries' valuations of abatement is rather ambiguous. El-Sayed and Rubio (2014) present a model where the development of cleaner technologies reduces a country's BAU emissions, so that for the same level of emissions, abatement costs are lower. The rest of the model is very similar to the one analyzed in this paper, but the study does not focus on the development of a breakthrough technology.

On the other hand, Ruis and de Zeeuw (2010), Urpelainen (2014) and Battaglini and

Harstad (2015) study the case of climate negotiations that include the possibility of cooperating in the development of a cleaner technology. Battaglini and Harstad (2015) use a dynamic game to include the duration of the agreement at the negotiation stage. Their findings show that if an incomplete agreement on emissions is signed, countries face a hold-up problem every time they negotiate. However, the free-rider problem can be mitigated and significant participation is feasible. Participation becomes attractive because only large coalitions commit to long-term agreements that avoid the hold-up problem. The hold-up problem has also been analyzed by Hong and Karp (2012) and Helm and Schmidt (2015) assuming that countries decide on R&D investment non-cooperatively before negotiating an abatement agreement. Hong and Karp (2012) show that with mixed strategies in the participation game, the small coalition paradox is overcome: equilibrium participation and welfare are higher in equilibria that involve higher investment. Mixed strategies create endogenous risk so that risk aversion increases the equilibrium probability of participation. However, Helm and Schmidt (2015) find the standard result that early R&D investments render free-riding more attractive. Finally, Goeschl and Perino (2015) find that the participation in an IEA can be small and the diffusion of green technology can reduce global abatement under trade-related intellectual property rights.[5]

To conclude the review of the literature, we would like to point out that the results obtained from the empirical papers are not conclusive. On the one hand, Buchner and Carraro (2004), Kemfert (2004) and Lessman and Edenhofer (2011) give support to the idea that supplementing an emissions agreement with technology elements or replacing it with a technology agreement can have positive effects on membership. However, Nagashima and Dellink (2008) and Nagashima et al. (2011) derive more pessimistic results. A final paper that deserves to be quoted is van der Pol et al. (2012), where the authors examine the effects of altruism on coalition stability. They distinguish between community and impartial altruism and find that community altruism, i.e. the exclusion

[5] Another strand of the literature studies the relationship between R&D investment in cleaner technologies and international cooperation but assuming that participation is exogenously given (see Buchholz and Konrad (1994), Beccherle and Tirole (2011), Harstad (2012, 2015) and Schmidt and Strausz (2014)). These papers focus on the negative effects of the hold-up problem.

of non-signatories from the concerns of signatories, can be quite effective to promote participation in an IEA.

To end this introduction we would like to highlight one feature of our approach that differentiates it from those used in the previous literature: we do not use an explicit abatement costs function in the stability analysis of the technology agreement. In fact, emissions abatement is not a variable in our model. Countries decide on R&D investment and energy production, and the level of emissions is determined by these two variables. Thus, the abatement costs are the investment costs the country has to support to reduce emissions when developing a cleaner technology and/or the reduction in benefit the country has to pay to reduce emissions by decreasing energy production. No other abatement costs are taken into account. Moreover, we also consider the technology spillovers. Such an approach can only be found in the analysis developed by Carraro and Siniscalco (1997), El-Sayed and Rubio (2014) and Schmidt and Strausz (2014).

The paper is organized as follows. The next section specifies the model. In Section 3, a technology agreement formation game is solved. This section also includes the analysis of the technology agreement with full participation and the analysis of the fully non-cooperative equilibrium. The stability analysis is developed in Subsection 3.3. Section 4 displays a numerical example that illustrates the main result of the paper: the stability of the grand coalition for large marginal damages. The conclusions drawn from this research are detailed in Section 5.

# 2 The model

We present a static model with $N$ countries that pollute the atmosphere and negotiate the control of greenhouse gas (GHG) emissions. Each country $i$ benefits from consuming energy $q_i,\ i = 1, ..., N$. As in much of the literature, we assume the benefit of consumption is represented by a linear-quadratic function:

$$B_i(q_i) = aq_i - \frac{q_i^2}{2}, \quad a > 0. \tag{1}$$

While energy consumption is privately beneficial, its production and consumption generates GHG emissions. The emission level of country $i$ is given by

$$e_i = (\theta - y_i)q_i, \quad \theta > 0, \tag{2}$$

where $y_i$ stands for the effective investment in cleaner technologies of country $i$ and $\theta - y_i$ is the emission intensity coefficient or the emissions-production ratio of the country's $i$ energy system.[6] With this specification of the emission function we are assuming that it is feasible to completely eliminate emissions with positive energy consumption if enough resources are invested in R&D. Then, we will say that a country has developed a breakthrough or clean technology if $y_i = \theta$ and that it uses a polluting or dirty technology if $y_i < \theta$.[7]

The effective investment of country $i$ depends on the amount invested in R&D in that country, $x_i$, and also on the investments in R&D undertaken in all other countries. However, technological diffusion is not perfect, only part of the R&D investments undertaken in other countries is beneficial for country $i$. Hence, the effective investment of country $i$ is given by

$$y_i = x_i + \gamma_i X_{-i}, \ \gamma \in (0, 1), \tag{3}$$

where $\gamma_i$ measures the degree of the spillover effects and $X_{-i} = \sum_{j \neq i} x_j$. Moreover, countries can reach larger technological spillovers by means of appropriate instruments of technological cooperation. Cooperating countries can allow for patents agreements that provide the other countries in the coalition with a large share of their own innovative technology. They can also sign agreements on technology transfers and/or joint R&D projects that increase the degree of innovation spillovers inside the coalition. Following the approach adopted by Kamien at al. (1992) in their analysis of the effects of R&D cartelization and research joint ventures (RJV) on oligopolistic competition, it is assumed that when countries cooperate they pool their R&D efforts so as to fully internalize

[6] This specification of the emissions function has been used, among others, by Carraro and Siniscalco (1997) and more recently by Ulph and Ulph (2007). Benchekroun and Chaudhuri (2014) assume that the emissions intensity coefficient is exogenously given and study how exogenous changes in technology affect the stability of an IEA.

[7] To simplify the notation we normalize the model doing $\theta = 1$, which implies that if effective investment is zero, then the level of emissions is equal to the level of energy production.

spillover effects, which implies that between signatories the degree of spillovers is equal to the unity and in the other cases is equal to $\gamma \in (0,1)$. Following the literature on R&D cooperation, $\gamma$ defines the "technological leakage". Hence $1-\gamma$ is the "differential technological leakage" between the signatories and the non-signatories or the "coalition information exchange".[8] Thus, if $n$ stands for the number of signatories, $s$ for a signatory country and $f$ for a non-signatory, the signatories' effective investment is

$$y_j^s = X^s + \gamma X^f = \sum_{k=1}^{n} x_k^s + \gamma \left( \sum_{l=1}^{N-n} x_l^f \right), \; j = 1, ..., n, \tag{4}$$

whereas the effective investment for non-signatories is given by (3).

Investment costs are quadratic: $C(x_i) = cx_i^2/2$, $c > 0$.

Finally, the environmental damages of country $i$ depend on global emissions:

$$E = \sum_{i=1}^{N} e_i = \sum_{i=1}^{N} (\theta - y_i) q_i, \tag{5}$$

and are assumed to be linear: $D(E) = dE$, $d > 0$. Besides, we impose that $2c < a^2$, a condition that guarantees that the signatories' effective investment of a technology agreement with full participation increases with marginal damages, $d$, and other results of the paper.[9] Thus, the net benefit of energy consumption for the representative country can be written as follows

$$W_i = aq_i - \frac{q_i^2}{2} - dE - \frac{c}{2} x_i^2, \tag{6}$$

where $E$ is given by (5), and $y_i$ is given by (3) for the non-signatories and by (4) for the signatories.

[8] Our specification of the spillovers is placed between two well-known cases: the (pure) *public good* and the *club good.* For a public good $\gamma = 1$ and then the "differential technological leakage" is zero. For a club good $\gamma = 0$ and the "differential technological leakage" is the unity.

[9] Notice that this constraint on parameter values is not very restrictive. The parameter $c$, that determines how steep the marginal investment costs are, could be larger than the maximum marginal benefit that the country derives from energy consumption, $a$. In other words, the constraint is satisfied if the slope of the marginal costs of investment is *relatively* low or, from the other side of the inequality, if the marginal benefit of energy is *relatively* large.

# 3 A technology agreement formation game

The formation of an international technology agreement is modeled as a three-stage game. Each stage will be described briefly in reverse order as the subgame-perfect equilibrium is computed by backward induction.

Given the level of membership and the investment in R&D of all countries, in the third stage, the emission game, each country simultaneously selects its national emissions acting unilaterally and taking the emissions of all other countries as given. If countries have developed a clean technology in the second stage, emissions will be zero in the third stage. In the second stage, the R&D investment game, the signatories select their R&D efforts so as to maximize the agreement's net benefit taking as given the R&D investments of non-signatories and pool their R&D investments so as to fully internalize the spillover effects inside the agreement. The non-signatories choose their investment in R&D acting non-cooperatively and taking the investments of all other countries as given in order to maximize their national net benefit. Both signatories and non-signatories choose their R&D investments simultaneously. Thus, R&D investments are provided by the *partial agreement Nash equilibrium* (PANE) with respect to a coalition defined by Chander and Tulkens (1995). Finally, it is assumed that in the first stage countries play a *simultaneous open membership game with a single binding agreement.* In a single agreement formation game, the strategies for each country are to sign or not to sign. The agreement is formed by all countries who have chosen to sign. Under open membership, any country is free to join the agreement if it is interested. Finally, we assume that the signing of the agreement acts as a binding device on the signatories. The game finishes when the emission subgame is over.

## 3.1 The third stage: an equilibrium in dominant strategies

As we have supposed that there is no cooperation in the third stage, optimal energy production can be calculated by maximizing the following net benefit function with respect to energy production

$$\max_{\{q_i\}} W_i = aq_i - \frac{q_i^2}{2} - d\left(\sum_{j=1}^{N}(1-y_j)q_j\right), \quad i = 1, ..., N$$

given that participation is decided in the first-stage and investments in the second stage.

The first-order condition for an *interior solution* is

$$a - q_i = d(1 - y_i), \tag{7}$$

where the left-hand side represents marginal benefit and the right-hand side *national* marginal damages of energy production. As there is no cooperation in this stage, each country only takes into account the consequences of its action on its own environmental damages. Notice that the national marginal damages of energy production are equal to the marginal damages of emissions times the emissions intensity of the energy system.

Thus, energy production is given by

$$q_i(y_i) = a - d(1 - y_i). \tag{8}$$

With linear environmental damages, the reaction functions of countries are orthogonal and optimal energy production levels are given by an equilibrium in dominant strategies. It is immediate from the above expression that energy production increases with effective investment.

Then, emissions are given by the following expression

$$e_i(y_i) = (1 - y_i)q_i(y_i) = (1 - y_i)(a - d(1 - y_i)). \tag{9}$$

At this point, we assume that $d < a/2$, which guarantees that effective investment reduces emissions.

Finally, net benefits can be written as a function of effective investments

$$W_i = aq_i(y_i) - \frac{1}{2}q_i(y_i)^2 - d\left(\sum_{j=1}^{N} e_j(y_j)\right) - \frac{c}{2}x_i^2, \quad i = 1, ..., N, \tag{10}$$

where energy production, $q_i(y_i)$, is given by (8) and emissions, $e_j(y_j)$, by (9).

## 3.2 The second stage: the PANE of the investment game

In this subsection, we solve stage two assuming that in the first stage $n$ countries, with $n \geq 2$, have signed the agreement. In this stage, non-signatories choose their investment in R&D acting non-cooperatively and taking the investments of all other countries, signatories and non-signatories, as given in order to maximize their national net benefit:

$$\max_{\{x_i^f\}} W_i^f = aq_i^f(y_i^f) - \frac{1}{2}q_i^f(y_i^f)^2 - d\left(\sum_{j=1}^{n} e_j^s(y_j^s) + \sum_{k=1}^{N-n} e_k^f(y_k^f)\right) - \frac{c}{2}(x_i^f)^2. \tag{11}$$

$$\text{s.t. } y_i^f = x_i^f + \gamma\left(\sum_{j=1}^{n} x_j^s + \sum_{k=1}^{N-n-1} x_k^f\right) \leq 1, \tag{12}$$

$$x_i^f \geq 0, \tag{13}$$

where constraint (12) defines the non-signatories' effective investment and establishes an upper bound for effective investment equal to the level of effective investment that implements a breakthrough technology. Moreover, the constraint guarantees that non-signatories' emissions are non-negative. Finally, (13) is the standard non-negative constraint on the decision variable, the R&D investment.

On the other hand, signatory countries coordinate their R&D efforts so as to maximize the net benefits of the international technology agreement (ITA) taking as given the R&D investments of non-signatories:

$$\max_{\{x_j^s\}} W_{ITA} = \sum_{l=1}^{n} W_l^s = \sum_{l=1}^{n}\left(aq_l^s(y_l^s) - \frac{1}{2}q_l^s(y_l^s)^2 - d\left(\sum_{l=1}^{n} e_l^s(y_l^s) + \sum_{k=1}^{N-n} e_k^f(y_k^f)\right) - \frac{c}{2}(x_l^s)^2\right) \tag{14}$$

$$= \sum_{l=1}^{n}\left(aq_l^s(y_l^s) - \frac{1}{2}q_l^s(y_l^s)^2 - \frac{c}{2}(x_l^s)^2\right) - nd\left(\sum_{l=1}^{n} e_l^s(y_l^s) + \sum_{k=1}^{N-n} e_k^f(y_k^f)\right), \quad j = 1, ..., n,$$

$$\text{s.t. } y_j^s = \sum_{l=1}^{n} x_l^s + \gamma \sum_{k=1}^{N-n} x_k^f \leq 1, \quad j = 1, ..., n, \tag{15}$$

$$x_j^s \geq 0, \quad j = 1, ..., n. \tag{16}$$

Constraints (15) and (16) have the same interpretation and play the same role as the constraints that appear in the non-signatories' maximization problem.

To derive the Kuhn-Tucker conditions that characterize the solutions of these non-linear programming problems, firstly we need to write the Lagrangian functions. For non-signatories, the Lagrangian function is

$$\mathcal{L}_i^f = W_i^f + \lambda_i^f (1 - y_i^f),$$

where $\lambda_i^f$ is the Lagrange's multiplier, and the Kuhn-Tucker conditions for the non-signatory representative country are

$$\frac{\partial \mathcal{L}_i^f}{\partial x_i^f} = a \frac{dq_i^f}{dy_i^f} \frac{\partial y_i^f}{\partial x_i^f} - q_i^f (y_i^f) \frac{dq_i^f}{dy_i^f} \frac{\partial y_i^f}{\partial x_i^f} - d \left( \sum_{j=1}^{n} \frac{de_j^s}{dy_j^s} \frac{\partial y_j^s}{\partial x_i^f} + \sum_{k=1}^{N-n} \frac{de_k^f}{dy_k^f} \frac{\partial y_k^f}{\partial x_i^f} \right) - c x_i^f - \lambda_i^f \frac{\partial y_i^f}{\partial x_i^f} \leq 0,$$

$$x_i^f \geq 0, \quad x_i^f \frac{\partial \mathcal{L}_i^f}{\partial x_i^f} = 0, \tag{17}$$

$$\frac{\partial \mathcal{L}_i^f}{\partial \lambda_i^f} = 1 - y_i^f \geq 0, \quad \lambda_i^f \geq 0, \quad \lambda_i^f \frac{\partial \mathcal{L}_i^f}{\partial \lambda_i^f} = 0. \tag{18}$$

According to (12) and (15)

$$\frac{\partial y_i^f}{\partial x_i^f} = 1, \ \frac{\partial y_j^s}{\partial x_i^f} = \frac{\partial y_k^f}{\partial x_i^f} = \gamma,$$

and the condition (17) for an interior solution can be rewritten as follows

$$\left( a - q_i^f (y_i^f) \right) \frac{dq_i^f}{dy_i^f} - d \left( \gamma \left( \sum_{j=1}^{n} \frac{de_j^s}{dy_j^s} + \sum_{k=1}^{N-n-1} \frac{de_k^f}{dy_k^f} \right) + \frac{de_i^f}{dy_i^f} \right) = c x_i^f, \tag{19}$$

where the first term of the left-hand side of the condition represents the increase in benefits coming from the increase in energy production caused by one more unit of R&D investment, and the second term stands for the reduction in environmental damages as a result of the reduction in global emissions caused by one more unit of R&D investment. Observe that the reduction in emissions from the rest of countries is discounted by $\gamma$, the degree of spillovers. On the right-hand side of the condition the marginal costs of the investment appear.

For signatories, the Lagrangian function is

$$\mathcal{L}_{ITA} = W_{ITA} + \sum_{j=1}^{n} \lambda_j^s (1 - y_j^s),$$

where $\lambda_j^s,\ j=1,...,n,$ are the Lagrange's multipliers and the Kuhn-Tucker conditions are

$$\frac{\partial \mathcal{L}_{ITA}}{\partial x_j^s}=\sum_{l=1}^{n}\left(a\frac{dq_l^s}{dy_l^s}\frac{\partial y_l^s}{\partial x_j^s}-q_l^s(y_l^s)\frac{dq_l^s}{dy_l^s}\frac{\partial y_l^s}{\partial x_j^s}\right)-nd\left(\sum_{l=1}^{n}\frac{de_l^s}{dy_l^s}\frac{\partial y_l^s}{\partial x_j^s}+\sum_{k=1}^{N-n}\frac{de_k^f}{dy_k^f}\frac{\partial y_k^f}{\partial x_j^s}\right)-cx_j^s$$

$$-\sum_{l=1}^{n}\lambda_l^s\frac{\partial y_l^s}{\partial x_j^s}\leq 0,\quad x_j^s\geq 0,\quad x_j^s\frac{\partial \mathcal{L}_{ITA}}{\partial x_j^s}=0,\quad j=1,...,n, \tag{20}$$

$$\frac{\partial \mathcal{L}_{ITA}}{\partial x_j^s}=1-y_j^s\geq 0,\quad \lambda_j^s\geq 0,\quad \lambda_j^s\frac{\partial \mathcal{L}_{ITA}}{\partial x_j^s}=0,\quad j=1,...,n. \tag{21}$$

According to (12) and (15)

$$\frac{\partial y_l^s}{\partial x_j^s}=\frac{\partial y_j^s}{\partial x_j^s}=1,\ \frac{\partial y_k^f}{\partial x_j^s}=\gamma,$$

and the condition (20) for an interior solution can be rewritten as follows

$$\sum_{l=1}^{n}\left(a-q_l^s(y_l^s)\right)\frac{dq_l^s}{dy_l^s}-nd\left(\sum_{l=1}^{n}\frac{de_l^s}{dy_l^s}+\gamma\sum_{k=1}^{N-n}\frac{de_k^f}{dy_k^f}\right)=cx_j^s. \tag{22}$$

If we compare this condition with that which characterizes the non-signatories' optimal investment, three differences can be observed. First, signatories take into account the positive effect that one more unit of R&D investment has on the benefits of the rest of signatories, not only on its national benefits as happens with non-signatories. Second, signatories also take into account the reduction in environmental damages for the rest of signatories caused by the reduction on global emissions of one more unit of R&D investment. Finally, notice that the reduction in emissions is discounted by $\gamma$, but only for non-signatories.

### 3.2.1 The technology agreement with full participation

We shall begin the study of the investment game solution addressing the case: $n=N$. In particular, we are interested in knowing under what conditions a technology agreement with full participation will develop a clean technology. With full participation in the agreement, condition (22) is[10]

$$N\left(a-q^s(y^s)\right)\frac{dq^s}{dy^s}-N^2d\frac{de^s}{dy^s}=cx^s,$$

[10] With ex-ante identical countries if all the countries participate in the agreement we expect a symmetric solution and the subscripts are not necessary.

which using (8) and (9) yields

$$Nd^2(1-y^s)+N^2d(a-2d(1-y^s))=cx^s.$$

Moreover, with full participation $y^s = Nx^s$ so that the effective investment for this case is

$$y^s=\frac{N^2d(aN-(2N-1)d)}{c-(2N-1)N^2d^2}. \tag{23}$$

It is easy to show that the numerator is positive if the effective investment decreases emissions, i.e. if $d < a/2$. Moreover, if $2c < a^2$ the critical value of $d$ for which the denominator is zero, $d_d^s = \left(c/(2N-1)N^2\right)^{1/2}$, is lower than $a/2$ and the effective investment *increases* with marginal damages. Then, since $y^s = 0$ for $d = 0$ and $\lim_{d\to d_d^s} y^s = +\infty$, a unique value for marginal damages exists

$$d_N^s=\frac{c}{aN^3}, \tag{24}$$

for which $y^s(d_N^s) = 1$. This means that the agreement develops a clean technology.

The following proposition summarizes this result.

**Proposition 1** *There exists a threshold value for marginal damages, $d_N^s = c/aN^3$, lower than $a/2$ such that if $d < d_N^s$, the effective investment is lower than 1 and increases with marginal damages. However, if $d \geq d_N^s$ the effective investment is equal to 1 and the technology agreement with full participation develops a breakthrough technology that eliminates emissions completely.*

**Proof.** See Appendix. ■

Next, we look at the other polar case: the fully non-cooperative equilibrium.

### 3.2.2 The fully non-cooperative equilibrium

We continue the study of the investment game solution calculating countries' effective investment when there is no cooperation. We are interested in knowing whether under this circumstance countries will develop a clean technology. If countries do not sign a technology agreement, condition (19) becomes:

$$\left(a-q^f(y^f)\right)\frac{dq^f}{dy^f}-d\left(\gamma\left(N-1\right)+1\right)\frac{de^f}{dy^f}=cx^f,$$

which using (8) and (9) yields

$$d^2(1-y^f)+d\left(\gamma\left(N-1\right)+1\right)\left(a-2d(1-y^f)\right)=cx^f.$$

On the other hand, if there is no cooperation $y^f = x^f(1+\gamma(N-1))$. Solving the system formed by these two equations we obtain the following expression for the effective investment:

$$y^f = \frac{d(1+\gamma(N-1))(a(1+\gamma(N-1))-d(1+2\gamma(N-1)))}{c-(1+\gamma(N-1))(1+2\gamma(N-1))d^2}. \tag{25}$$

This expression has identical features to those of expression (23), which defines the effective investment when all the countries have signed a technology agreement. The numerator is positive if $d < a/2$. On the other hand, if $2c < a^2$ the effective investment is increasing with respect to marginal damages and moreover if the spillover effects are not very low, the critical value of $d$ that makes the denominator zero, $d_d^f = (c/(1+\gamma(N-1))(1+2\gamma(N-1)))^{1/2}$, is lower than $a/2$. Then, since $y^f = 0$ for $d = 0$ and $\lim_{d\to d_d^f} y^f = +\infty$, there exists a unique value for marginal damages:

$$d_0^f = \frac{c}{a(1+\gamma(N-1))^2}, \tag{26}$$

which satisfies $y^f(d^f) = 1$. If marginal damages are large enough, countries unilaterally implement a clean technology.

Summarizing,[11]

**Proposition 2** *If the spillover effects are not very low, there exists a threshold value for marginal damages, $d^f = c/a(1+\gamma(N-1))^2$, lower than $a/2$ such that if $d < d_0^f$, the effective investment is lower than 1 and increases with marginal damages. However, if $d \geq d_0^f$, the effective investment is equal to 1 and countries develop non-cooperatively a breakthrough technology that completely eliminates emissions.*

[11]We do not present the details of the proof in this case because they follow step by step those of the proof for proposition 1. The only difference is that now $2c < a^2$ does not guarantee that the proposition holds if the spillover effects are very low.

This result establishes that cooperation is not strictly necessary to implement a breakthrough technology. If marginal damages are large enough, countries acting unilaterally find it profitable to invest in R&D to eliminate national emissions. However, the optimal level of investment is not implemented because countries do not take into account the positive externalities that investment has on other countries. Next, we show that effective investment is lower when countries do not sign a technology agreement.

First, we compare the threshold values of marginal damages for each solution.

$$d_0^f - d_N^s = \frac{c}{a(1+\gamma(N-1))^2} - \frac{c}{aN^3} = \frac{c}{a}\frac{N^3 - (1+\gamma(N-1))^2}{(1+\gamma(N-1))^2 N^3} > 0 \text{ for } \gamma \in (0,1).$$

Thus, we can conclude that

**Lemma 1** *A technology agreement with full participation completely eliminates emissions for a level of marginal damages lower than the level for which countries eliminate emissions unilaterally, i.e.* $d_N^s < d_0^f$.

Now, using (23), (25) and this lemma, the following result can be derived

**Proposition 3** *The effective investment of the fully non-cooperative equilibrium is lower than the effective investment selected by the signatories of a technology agreement with full participation, i.e.* $y^f < y^s$ *for* $d \in (0, d_0^f)$ *with* $y^s = 1$ *from* $d \geq d_N^s$ *and* $y^f = y^s = 1$*when* $d = d_0^f$.

**Proof.** See Appendix. ■

## 3.3 The first stage: the stability of the grand coalition

In this subsection, we investigate whether a technology agreement consisting of all countries developing a clean technology is stable. First, we present the concept of coalitional stability used in this paper that was proposed by d'Aspremont et al. (1983) and has been extensively applied in the literature on IEAs.

**Definition 1** *An agreement consisting of* $n$ *signatories is stable if* $W_j^s(n) \geq W_j^f(n-1)$ *for* $j = 1, ..., n$ *and* $W_i^f(n) \geq W_i^s(n+1)$ *for* $i = 1, ..., N-n$.

The first inequality, which is also known as the *internal stability condition*, simply means that any signatory country is at least as well-off staying in the agreement as withdrawing from it, assuming that all other countries do not change their membership status. The second inequality, which is also known as the *external stability condition*, similarly requires any non-signatory to be at least as well-off remaining a non-signatory as joining the agreement, assuming once again, that all other countries do not change their membership status.

According to this concept of coalitional stability, if we focus on the stability of the grand coalition, we only need to check if the internal stability condition is satisfied, i.e. to check if $W_j^s(N) \geq W_j^f(N-1)$. Moreover, according to Proposition 1 if we are interested in investigating the stability of the grand coalition when it develops a breakthrough technology, we have to assume that $d \geq d_N^s$. In that case, the energy production of the signatories is $a$, the emissions are zero, the R&D investment is $1/N$ and the signatories achieve the following level of net benefit

$$W^s(N) = \frac{a^2}{2} - \frac{c}{2N^2}. \tag{27}$$

To calculate $W_j^f(N-1)$ we need to know those investments selected by the signatories and the non-signatories when the technology agreement is signed by $N-1$ countries. Firstly, we study the possibility that both the non-signatory and the signatories develop a clean technology. Using the Kuhn-Tucker conditions we can find out under which conditions this could occur. In particular, the Kuhn-Tucker conditions allows us to calculate a threshold value for marginal damages that separates the corner solution (with both types of countries eliminating completely emissions) from the interior solution (with at least the non-signatory selecting a level of effective investment lower than the unity).

The Kuhn-Tucker conditions defined above imply that the following inequalities must be satisfied when the participation is equal to $n$

$$(1+\gamma(N-1))ad - (1+2(N-n-1)\gamma)d^2(1-y^f) - 2n\gamma d^2(1-y^s) - cx^f - \lambda^f \leq 0,$$

$$1 \geq y^f = x^f(1+\gamma(N-n-1)) + \gamma n x^s,$$

$$n(n+(N-n)\gamma)ad-(2n-1)nd^2(1-y^s)-2(N-n)n\gamma d^2(1-y^f)-cx^s-n\lambda^s \leq 0,$$

$$1 \geq y^s = nx^s + \gamma(N-n)x^f.$$

For $n = N-1,\ \lambda^f = 0,\ y^f = y^s = 1$, we can calculate the threshold value for marginal damages that supports this corner solution solving the following system of equations

$$(1+\gamma(N-1))ad - cx^f = 0, \quad (28)$$

$$1 - x^f - \gamma(N-1)x^s = 0, \quad (29)$$

$$(N-1)(N-1+\gamma)ad - cx^s - (N-1)\lambda^s = 0, \quad (30)$$

$$1-(N-1)x^s - \gamma x^f = 0. \quad (31)$$

From (29) and (31) we obtain the effective investment for both the signatories and the non-signatory.

$$0 < x^s(N-1) = \frac{1}{(N-1)(1+\gamma)} < x^f(N-1) = \frac{1}{1+\gamma}.$$

Observe that the signatories' R&D investment is lower than the effort in R&D undertaken by the non-signatory. This results from the fact that signatories pool their R&D efforts so as to fully internalize spillover effects and that both signatories and the non-signatory invest to reach the same level of effective investment.

Substituting $x^f$ in (28) we obtain a lower bound for the marginal damages above which both types of countries develop a clean technology. This threshold value is

$$d^f_{N-1} = \frac{c}{a(1+\gamma)(1+\gamma(N-1))}. \quad (32)$$

It is easy to show, as it could be expected, that $d^f_0 < d^f_{N-1} < a/2$.[12]

This analysis can be summarized as follows

**Lemma 2** *There exists a threshold value for marginal damages,* $d^f_{N-1} \in (d^f_0, a/2)$*, such that if* $d \in [d^f_{N-1}, a/2)$ *then* $y^s(N-1) = y^f(N-1) = 1$.

[12] It is also easy to check that $\lambda^s$ is positive indicating that the signatories develop a clean technology for lower values of marginal damages.

Thus, for $d \geq d^f_{N-1}$, the energy production of the non-signatory is $a$, emissions are zero, the R&D investment is $1/(1+\gamma)$ and the level of net benefit is

$$W^f(N-1) = \frac{a}{2} - \frac{c}{2(1+\gamma)^2}. \tag{33}$$

Comparing this expression with (27), it is immediate that the internal stability condition holds. Thus, we can conclude that for $d \in (d^f_{N-1}, a/2)$, the signatories implement a clean technology that completely eliminates emissions and the grand coalition is stable.

When $d$ is lower than $d^f_{N-1}$, the non-signatory country's effective investment is lower than the unity, and so the country produces energy with a polluting technology but signatories could still use a clean technology. Applying the Kuhn-Tucker conditions once again, we can investigate under which conditions this could happen. Now the Kuhn-Tucker conditions allow us to derive the threshold value for marginal damages that separates the interior solution (with both types of countries developing a polluting technology) from the solution for which the non-signatory country still uses a polluting technology but the signatories develop a clean technology. For $n = N-1$, $\lambda^f = \lambda^s = 0$ and $y^s = 1$ in (28), (29), (30) and (31), the following system of equations is obtained

$$\begin{aligned}
(1+\gamma(N-1))ad - d^2(1-y^f) - cx^f &= 0, & (34)\\
y^f - x^f - \gamma(N-1)x^s &= 0, & (35)\\
(N-1)(N-1+\gamma)ad - 2(N-1)\gamma d^2(1-y^f) - cx^s &= 0, & (36)\\
1-(N-1)x^s - \gamma x^f &= 0. & (37)
\end{aligned}$$

Solving (34), (35) and (37) we obtain the R&D investments and the non-signatory's effective investment:

$$\begin{aligned}
x^s(N-1) &= \frac{1}{N-1}\left(\frac{c - d^2(1-\gamma) - \gamma ad(1+\gamma(N-1))}{c-d^2(1-\gamma^2)}\right), & (38)\\
x^f(N-1) &= \frac{d(a(1+\gamma(N-1)) - d(1-\gamma))}{c-d^2(1-\gamma^2)}, & (39)\\
y^f(N-1) &= \frac{d(1-\gamma^2)(a(1+\gamma(N-1))-d)+c\gamma}{c-d^2(1-\gamma^2)}, & (40)
\end{aligned}$$

and eliminating by substitution $x^s$ and $y^f$ in (36), the following equation in $d$ is obtained:

$$(N-1)^2(1-\gamma^2)(2(N-1)\gamma^2+\gamma-(N-1))ad^3 - (1-\gamma)(2(N-1)^2\gamma-1)cd^2$$

$$+((N-1)\gamma^2+(N^2-2N+2)\gamma+(N-1)^3)acd-c^2=0, \qquad (41)$$

where the first coefficient of the left-hand side of the equation is zero for the combinations $(\gamma, N)$ defined by

$$\gamma_1(N)=\frac{-1+(1+8(N-1)^2)^{1/2}}{4(N-1)} \in [0.5, 0.71) \text{ for all } N \geq 2,$$

and the second coefficient is zero for the combinations $(\gamma, N)$ defined by

$$\gamma_2(N)=\frac{1}{2(N-1)^2}<0.5 \text{ if } N>2,$$

with

$$\gamma_2(N)=\frac{1}{2(N-1)^2}<\gamma_1(N)=\frac{-1+(1+8(N-1)^2)^{1/1}}{4(N-1)} \text{ for } N>2.$$

Then, according to the Descartes' rule of signs the polynomial equation (41) could have a maximum of two positive real roots if $\gamma \in (0, \gamma_1(N)]$ and a maximum of three positive real roots if $\gamma \in (\gamma_1(N), 1)$.[13] However, the Descartes' rule of signs only establishes an upper bound in the number of positive roots the equation may have. Next, we show that the equation has only one positive root lower than $d^f_{N-1}$ which will allow us to conclude that there exists a threshold value for marginal damages above which the signatories completely eliminate emissions but the non-signatory country develops a polluting technology. With this aim, we substitute $d$ for $d^f_{N-1}$ in (41). The substitution leads to the following expression

$$\frac{c^3(1-\gamma)}{(1+\gamma)(1+\gamma(N-1))^3a^2}((1+\gamma)(1+\gamma(N-1))^2(N-2)(\gamma(N-1)+N^2-N+1)\frac{a^2}{c}$$
$$-(1-\gamma)((N-1)(N-2)\gamma+(N-1)^3-1)),$$

which is positive if

$$\frac{a^2}{2c}>\frac{(1-\gamma)((N-1)(N-2)\gamma+(N-1)^3-1)}{2(1+\gamma)(1+\gamma(N-1))^2(N-2)(\gamma(N-1)+N^2-N+1)}.$$

[13]If $\gamma \in (0, \gamma_2(N)]$, the first coefficient is negative, the second one is positive or zero, the third one is positive and the independent term is negative. There are two changes in the sign of the coefficients. If $\gamma \in (\gamma_2(N), \gamma_1(N)]$, the first coefficient is negative or zero, the second one is negative, the third one is positive and the independent tern is negative. Again there are two changes in the sign of the coefficients. Finally, if $\gamma \in (\gamma_1(N), 1)$, the first coefficient is positive, the second one is negative, the third one is positive and the independent term is negative and in this case there are three changes in the sign of the coefficients.

As we have assumed that $a^2 > 2c$, this condition is satisfied if

$$1 - \frac{(1-\gamma)((N-1)(N-2)\gamma + (N-1)^3 - 1)}{2(1+\gamma)(1+\gamma(N-1))^2(N-2)(\gamma(N-1) + N^2 - N + 1)} > 0.$$

The numerator of this difference is

$$2(N-2)(N-1)^3\gamma^4 + 2\left(N^2+2\right)(N-2)(N-1)^2\gamma^3 + (N-2)(N-1)\left(4N + 2N^3 + 1\right)\gamma^2$$
$$+ (N-2)\left(6N - 5N^2 + 4N^3 - 2\right)\gamma + (N-2)(N^2 - N + 1)$$

which is positive for $\gamma \in (0,1)$ and $N \geq 2$. Thus, we find that the left-hand side of equation (41) is positive for $d = d^f_{N-1}$.

Then, if the polynomial on the left-hand side of equation (41) is positive for $d = f^f_{N-1}$, negative for $d = 0$ and strictly concave and decreasing for large values of $d$ which requires that $\gamma \in (0, \gamma_1(N)]$, we can conclude that the equation has two positive roots for $d$ and that $d^f_{N-1}$ belongs to the interior of the interval defined by these two positive roots.[14] In other words, for $\gamma \in (0, \gamma_1(N)]$ the equation (41) has two positive roots and that the lowest root is the threshold value for marginal damages, which we call $d^s_{N-1}$, and separates the interior solution with both types of countries developing a polluting technology from the solution for which the non-signatory country uses a polluting technology but the signatories do not.

Next, we check that investments are non-negative in the interval $[d^s_{N-1}, d^f_{N-1})$. Comparing the numerators of (38) and (39) and with the common denominator of both expressions, it is easy to show that investments are positive both for the signatories and the non-signatory whether the marginal damages are lower than the positive root that makes the numerator of the signatories' investment zero. Thus, if we substitute $d$ for $d^f_{N-1}$ in the numerator of (38) and find that the value of polynomial is positive, then we can conclude that investments are positive in the interval $[d^s_{N-1}, d^f_{N-1})$. The substitution yields the following expression

$$\frac{c(1-\gamma)}{(1+\gamma)(1+\gamma(N-1))^2a^2}((1+\gamma)(1+\gamma(N-1))^2a^2 - (1-\gamma)c),$$

[14] When $\gamma \in (\gamma_1(N), 1)$, the equation could have three positive roots, which makes the analysis of the existence and uniqueness of a solution more complicated. Nevertheless, we expect that also in this case, the equation (41) defines a unique threshold value for marginal damages. In fact, in the Appendix we show that for $\gamma \geq 0.71$ the equation (41) has a unique solution that is lower than $d^f_{N-1}$.

which is positive for $a^2 > 2c$. Thus, the investments are positive in the interval $[d^s_{N-1}, d^f_{N-1})$ and by definition the effective investment of the non-signatory is also positive. Finally, using the same procedure we have just applied to show that the value of the polynomial in the left-hand side of equation (41) is positive for $d^f_{N-1}$we can establish that $d^s_N < d^s_{N-1} < d^f_0$.

All the previous analysis can be summarized in the following result

**Lemma 3** *If $\gamma \in (0, \gamma_1(N)]$, then there exists a threshold value for marginal damages, $d^s_{N-1} \in (d^s_N, d^f_0)$, such that if $d \in [d^s_{N-1}, d^f_{N-1})$ then $y^s(N-1) = 1$ and $y^f(N-1) < 1$.*

Next, we study the stability of the grand coalition for values of marginal damages in the interval $[d^s_{N-1}, d^f_{N-1})$. For these values of marginal damages, the non-signatory develops a polluting technology which implies that its environmental damages are positive and that the energy production is lower than $a$, the level of energy production a country achieves when using a clean technology. This means that a country supports larger damages and enjoys a lower benefit if it leaves the grand coalition. Then if it also supports larger or the same investment costs as happened in the interval $[d^f_{N-1}, a/2)$, we could conclude that it is not profitable for the country to leave the grand coalition. In other words, the internal stability condition would be satisfied and the grand coalition would be stable.

Using (39) we can calculate the difference in the levels of investment:

$$x^f(N-1) - x^s(N) = \frac{d(a(1+\gamma(N-1)) - d(1-\gamma))}{c - d^2(1-\gamma^2)} - \frac{1}{N}$$

$$= \frac{(1-\gamma)(N-1-\gamma)d^2 - N(1+\gamma(N-1))ad + c}{N((1-\gamma^2)d^2 - c)}. \tag{42}$$

If $2c < a^2$ it is easy to check that the numerator of (42) is zero for two positive values of marginal damages and also establish the following ordering

$$0 < d^-_{nx} < d^f_0 < d^f_{N-1} < d_{dx} < d^+_{nx}, \tag{43}$$

where $d^-_{nx}$ and $d^+_{nx}$ are the values of marginal damages that make the numerator of the difference in the levels of investment zero and $d_{dx}$ is the value that makes the denominator zero. The ordering indicates that in the interval $(d^-_{nx}, d^f_{N-1})$, both the numerator and the denominator of (42) are negative and therefore the difference in the levels of investment is positive. When $d = d^-_{nx}$, the levels of investment are the same. Unfortunately, it is not possible to establish the relationship between

$$d^-_{nx} = \frac{N(1+\gamma(N-1))a - (N^2(1+\gamma(N-1))^2a^2 - 4(1-\gamma)(N-1-\gamma)c)^{1/2}}{2(1-\gamma)(N-1-\gamma)} \tag{44}$$

and $d^s_{N-1}$, which is implicitly defined by (41) although we conjecture that $d^s_{N-1}$ is lower than $d^-_{nx}$. If this is the case, we can conclude that for values of marginal damages in the interval $[d^-_{nx}, d^f_{N-1})$ , $W^s(N) > W^f(N-1)$. The internal stability condition for the grand coalition is satisfied and a technology agreement with full participation that develops a breakthrough technology is stable. If this is not the case and $d^s_{N-1}$ is larger than $d^-_{nx}$, we would arrive to the same conclusion but for values of marginal damages in the interval $[d^s_{N-1}, d^f_{N-1})$.

Summarizing, we have shown above that the grand coalition is stable for $d \in [d^f_{N-1}, a/2)$ and we have just concluded that the grand coalition is also stable in an interval with an upper extreme equal to $d^f_{N-1}$ and a lower extreme equal to the maximum value of the pair $\{ d^s_{N-1}, d^-_{nx}\}$. Thus, we can summarize the analysis of the stability of the grand coalition developed in this subsection as follows

**Proposition 4** *If* $\gamma \in (0, \gamma_1(N)]$ *and* $\max\{ d^s_{N-1}, d^-_{nx}\} \leq d < a/2$, *then the grand coalition develops a breakthrough technology and is stable.*

Nevertheless, we expect that this result holds for larger values of $\gamma$ and lower values of $d$. In fact, the argument we have just presented to check the stability of the grand coalition also applies if $\gamma \geq 0.71$ since we know that the equation (41) has a unique solution, $d^s_{N-1}$, lower than $d^f_{N-1}$ in this case. Then if the model is parsimonious, it is reasonable to think that the grand coalition be stable for values of $\gamma$ in the interval $(\gamma_1(N), 0.71)$ above all if it is taken into account that the difference $0.71 - \gamma_1(N)$ is small and decreasing with $N$. For instance, for $N = 10$ this difference is 0.0301 and for $N = 100$ it is 0.0054. Moreover, our analysis shows that spillovers play for cooperation provided that the degree of spillovers is less than the unity to avoid the investment becoming a (pure) public good.[15] The way that spillovers promote cooperation is through their influence in the threshold values $d^s_{N-1}$ and $d^-_{nx}$. In the appendix it is shown that both threshold values diminish with the degree of spillovers. According to the previous proposition, this increases the interval of values for marginal damages for which the grand coalition is stable. Thus, spillovers promote cooperation because the larger the spillover the lower the marginal damages that justify the development of a clean technology and stabilize the grand coalition.

On the other hand, it is pretty obvious that if $d^s_{N-1} < d^-_{nx}$ then for marginal damages close enough to $d^-_{nx}$ from the left, the grand coalition must be stable. Notice that for $d^-_{nx}$ the investment of a country when it is a non-signatory is the same as the investment that country selects as a member of the grand coalition. This means that for $d^-_{nx}$ the country's net benefit when it is a signatory of the grand coalition is larger than the country's net benefit when it is the only non-signatory of an agreement consisting of $N-1$ signatories, since in this second case it supports larger damages and enjoys a lower benefit. Thus, there must exist a critical value for marginal damages lower than $d^-_{nx}$ such that for values of marginal damages larger than this critical value and lower than $d^-_{nx}$, the grand coalition remains stable. In this interval, the investment costs are larger when the country is a signatory than when it is the only non-signatory but the reduction in investment costs is more than compensated by the increase in damages and the decrease in the benefit coming from the reduction in energy production. In the next section, we present a numerical example that illustrates this argument.

# 4 A numerical illustration

We shall begin calculating the threshold values for the marginal damages we have defined in the previous section for the following set of parameter values: $a = 10,\ N = 10,\ c =$

[15]Notice that if $\gamma = 1$ there is no difference in the effective investment between signatories and non-signatories and then Lemma 3 does not hold. In other words, it is not possible to have an equilibrium in the investment game for which the signatories completely eliminate emissions and the non-signatories develop a polluting technology.

20, $\theta = 1$ and $\gamma = 0.05$. The resulting values are[16]

$$d^s_{10} = 0.0020 < d^s_9 = 0.0027 < d^-_{nx} = 0.1391 < d^f_0 = 0.9512 < d^f_9 = 1.3136 < a/2 = 5.$$

$d^s_{10}$ stands for the minimum marginal damages that justify the grand coalition to develop a clean technology and $d^f_0$ for the minimum marginal damages for which countries develop a clean technology unilaterally, i.e. without cooperation. $d^s_9$ has the same interpretation as $d^s_{10}$ but for an agreement consisting of nine countries and $d^f_9$ are the minimum marginal damages that justify the non-signatory of an agreement consisting of nine countries to develop a clean technology. For values of marginal damages equal to or above $d^f_9$, both signatories and non-signatories completely eliminate emissions for all levels of cooperation. Finally, $d^-_{nx}$ is the threshold value for marginal damages for which the investment of the only non-signatory of an agreement formed by nine countries is equal to the signatories' investment of the grand coalition i.e. $x^f(9) = x^s(10)$. Above this threshold value the non-signatory's investment is larger than the signatories' investment of the grand coalition. In this case, if one country leaves the grand coalition, it would have to support larger investment costs, which would completely eliminate the incentives to leave the grand coalition. Thus, according to Proposition 4, in our example if $d \in [0.1391, 5)$, the grand coalition develops a breakthrough technology and is stable.

$\Rightarrow$ TABLE 1A $\Leftarrow$

Tables 1A and 1B show the PANE of the investment game for the different membership when $d = d^s_{10}$. For this value of marginal damages the solution of the game is an interior solution for the effective investment and emissions, except for $n = 10$. The effective investment for both signatories and non-signatories is lower than the unity, which

[16] Remember that we have normalized the model assuming that $\theta = 1$, which implies that if effective investment is zero the level of emissions is equal to the level of energy production. This assumption yields small values for the threshold values and hence yields levels of energy production that are close to $a$. Take into account that if marginal damages of emissions are low, the marginal damages of the energy production will be also low and $q$ will not be very far from $a$. See expression (7). Obviously, this is a consequence of the normalization we have assumed. For larger values of $\theta$, larger values for investment would be obtained and the gap between $a$ and $q$ would increase.

yields positive emissions i.e. both signatories and non-signatories produce energy using a polluting technology. Only when all countries sign up to the technology agreement is a clean technology developed. The properties of the solution are the usual properties for games with positive spillovers coming from cooperation. The investment, the effective investment and the energy production of signatories are larger than the levels selected by non-signatories. Signatories make a stronger effort to reduce pollution and consequently the energy production of signatories is larger than the energy production of non-signatories. Take into account that investment reduces the emissions intensity coefficient, which decreases the marginal damages of the energy production and thus leads to the country selecting a higher level of energy production. Moreover, all these variables increase with membership.

$\Rightarrow$ TABLE 1B $\Leftarrow$

In Table 1B, emissions and net benefits are shown. As usual, the signatories' emissions are lower than those of non-signatories, whereas the relationship is the contrary for the net benefit. Countries' emissions decrease with the participation in the agreement and net benefits increase. The last column displays one the most important characteristics of this type of game: the difference in net benefit increases with cooperation. This characteristic explains that the greater the participation, then the larger the incentives to leave the agreement become and therefore high levels of participation cannot be reached. In the example, the stable agreement consists of six countries.[17]

$\Rightarrow$ TABLE 2A $\Leftarrow$

In the second example shown in Tables 2A and 2B, $d$ is equal to $0.107 \in (d_9^s = 0.0027, d_{nx}^{-} = 0.1391)$. For $d = 0.107$, an agreement formed by three countries already implements a clean technology. Thus, for $n \geq 3$ the solution of the game is a corner solution

[17] This result is consistent with that obtained by El-Sayed and Rubio (2014) for a similar model. These authors show that when emissions are *positive* the membership cannot be larger than six countries regardless of the number of countries involved in the externality. The main difference between our model and that analyzed by El-Sayed and Rubio (2014) is that these authors do not explicitly take into account the decision on energy production.

for signatories' effective investment and emissions. The signatories' effective investment is equal to the unity and emissions are zero. The signatories produce energy using a clean technology whereas the non-signatories use a polluting technology. Nevertheless, the properties of the investment game are the same as those that characterize the previous example: the signatories' investment, effective investment and energy production are greater than the levels selected by the non-signatories whereas the signatories' emissions and net benefit are lower. However, the increase in membership has different effects on the variables of the model. Although the signatories' investment increases with the participation for low membership once the agreement is signed by three countries, the signatories' investment *decreases* with participation. This effect is a consequence of the fact that the signatories' effective investment reaches the maximum with three signatories so that any new incorporation to the agreement can only reduce the signatories' investment. A first consequence of this change is that non-signatories' effective investment decreases with membership, which also causes a reduction in non-signatories' energy production. Notice that if effective investment is lower, the marginal damages of energy production are larger and the country will select a lower level of energy production.

$\Rightarrow$ TABLE 2B $\Leftarrow$

The net effect of a reduction in the effective investment and energy production is that non-signatories' emissions increase with participation. This is possible because reduction in the effective investment increases the emissions intensity coefficient, so that non-signatories produce less energy but use a more polluting technology, resulting in larger emissions for the parameter values of the example. Nevertheless, cooperation reduces global emissions. The development of a clean technology by the signatories is enough to compensate for the increase of emissions caused by non-signatories. However, all these changes do not modify the effect of an increase in participation on the net benefits. These continue to increase with respect to the participation for both signatories and non-signatories although, as the last column of Table 2B shows, there is a change in the difference of net benefits. Now, this difference increases until the agreement consists of three countries and a clean technology is developed. From this value, the difference in net benefits decreases and the incentives to leave the coalition also decrease with participation. In the example, this effect is strong enough to stabilize the grand coalition. Observe that as $d$ is lower than $d_{nx}^{-}$, the investment of the non-signatory when the agreement consists of nine countries is lower than the signatories' investment in the grand coalition: $x^s(10) = 0.1 > x^f(9) = 0.07707$. Nevertheless, the grand coalition is stable. In this case, the reduction in investment costs is more than compensated by the increase in damages since the aggregate emissions are zero when all countries participate in the agreement but they are positive for an agreement consisting of nine countries: $E(10) = 0 < E(9) = 8.65$, and because of the reduction in the benefit caused by the decrease in energy production: $q^s(10) = 10 > q^f(9) = 9.9065$. The net effect is that the net benefit of a country belonging to the grand coalition is larger than the net benefit the country can obtain as a non-signatory: $W^s(10) = 49.90 > W^f(9) = 49.01$. The internal stability condition is satisfied and the grand coalition is stable. Thus, as we pointed out in the previous section, we should expect the grand coalition to be stable for lower values of marginal damages than those that appear in Proposition 4.

$\Rightarrow$ TABLE 3A $\Leftarrow$

Finally, we present an example for which $d = 0.25 \in (d_{nx}^{-} = 0.1391, d_0^f = 0.9512)$. For this example as $d$ is larger than $d_{nx}^{-}$ we obtain that $x^s(10) = 0.1 < x^f(9) = 0.1788$. With large enough environmental damages, the investment is greater when a country does not belong to the grand coalition. The features of this example coincide with those of the previous example except in one thing: the signatories' investment is larger than the non-signatories' investment for $n \leq 5$ but this relationship reverses for $n \geq 6$. With larger environmental damages, the non-signatories invest more for any level of cooperation as can be checked comparing the third column in Tables 2A and 3A. However, this only occurs for the signatories when $n = 2$. The comparison of the second column shows that the signatories' investment is larger for $n$ in the interval $[3, 9]$ when $d = 0.107$ than when $d = 0.25$. The fact that effective investment for signatories is bounded above for all $n \geq 2$ and that the spillovers coming from the non-signatories are greater because their investment is higher explains that the signatories invest less for higher environmental damages. Thus, the combination of lower and decreasing investments by the signatories with larger and increasing investments by the non-signatories account for the reversion in the relation between these two variables. This reversion in investment causes a reversion in the relationship between net benefits as the appearance of a negative sign in the last column of Table 3B indicates. More cooperation is still good for the non-signatories but because of the reduction in investment costs, for $n \geq 6$ the signatories' net benefit is greater than the non-signatories' net benefit. The grand coalition is stable as Proposition 4 establishes. For $d = 0.25$, if a country leaves the agreement it not only has to support larger environmental damages and a lower benefit because of the reduction in the energy production, as it happened for $d = 0107$, but also higher investment costs.

$\Rightarrow$ TABLE 3B $\Leftarrow$

These examples also indicate that we should expect that the minimum participation required to develop a clean technology falls with marginal damages. In the first example with $d = 0.002$, the clean technology cannot be developed if the agreement does not consists of all countries. However, in the second example with $d = 0.107$, the development of a clean technology by the signatories requires an agreement with a minimum membership of three countries, that reduces to two for $d = 0.25$.

As usual, the gains coming from cooperation growth with marginal damages. For $d = 0.002$, the grand coalition could only achieve a percentage increase in the net benefit of the countries around 0.2%, whereas for $d = 0.107$ the full cooperation increases the countries' net benefit by 23.2% and for $d = 0.25$ by 58.8%. In all these cases, the grand coalition develops a "breakthrough technology". Our analysis shows that in the last two cases when the gains from full cooperation are larger, the grand coalition is stable.

# 5 Conclusions

This paper analyzes whether a technology agreement might be a good alternative to an emissions agreement that fails to deliver significant emissions abatements because of the free-rider problem. The focus is on breakthrough technologies that completely eliminate emissions. Our analysis shows that the grand coalition is stable provided that marginal damages are great enough to make the development of a breakthrough technology profitable. Stability is achieved jointly with the development of a breakthrough technology. We find that there is a threshold value for marginal damages above which if a country leaves the grand coalition it has to support greater investment costs: so the exit becomes unprofitable and the grand coalition stable. Although we have derived a sufficient condition for the stability of the grand coalition, we expect that the grand coalition is stable for lower values of marginal damages that do not satisfy this sufficient condition. Our conjecture is based on a continuity argument. It seems clear that for values of marginal damages below (but close to) the threshold value, although the country leaving the agreement invests less, it will have to support larger damages and a lower benefit of consumption because of the reduction in energy production caused by exiting the agreement. Therefore, the net effect in this case could lead to a reduction in the country's net benefit, which would make leaving the agreement unprofitable. Notice that if the country develops a polluting technology, the marginal damages of the energy production become positive, since these marginal damages are equal to the marginal damages of emissions times the emissions intensity coefficient. Consequently the country leaving the agreement will reduce both its production and consumption of energy. In Section 4, we have presented a numerical example that illustrates this argument. Although our model is pretty stylized as to think that our conclusions might be definitive, our analysis indicates that there exists a potential for successful cooperation if the focus of climate negotiations is on the development of green technologies that allow countries to move away from fossil-fuels. Thus, the policy recommendation resulting from our findings suggests that an agreement to control climate change should settle the foundations of a new energy system definitively based on non-polluting technologies.

There are a lot of interesting extensions that might be addressed in the future to check the robustness of our results in detail. The more obvious extensions are to consider increasing marginal damages and heterogeneous countries. We do not expect qualitative changes if marginal damages are increasing. With heterogeneous countries the grand coalition might be unstable, if there are important differences in the investment costs but only when marginal damages are small enough to lead the country leaving the agreement to select a lower investment. Another interesting extension would be to investigate how the technology agreement could affect the competitiveness of signatories in international markets. Finally, the effects of uncertainty on both the success of investment and the severity of marginal damages are also key issues to take into account in the agenda for future research.

**Appendix**

**A) Proof of Proposition 1**

The sketch of the proof is the main text. Here, we present the details. Firstly, we calculate the difference

$$\left(\frac{a}{2}\right)^2 - (d_d^s)^2 = \frac{a^2}{4} - \frac{c}{(2N-1)N^2},$$

which is positive if

$$\frac{2}{(2N-1)N^2} < \frac{a^2}{2c}.$$

As we have assumed that $1 < a^2/2c$, this inequality is satisfied provided that $1-(2/(2N-1)N^2)$ is positive, which is true for $N \geq 2$. Then $d_d^s < (a/2)$.

On the other hand, the derivative of the effective investment with respect to $d$ is

$$\frac{\partial y^s}{\partial d} = \frac{N^2(aN^3(2N-1)d^2 - 2(2N-1)cd + acN)}{(c-(2N-1)N^2d^2)^2}. \tag{45}$$

The polynomial of second degree in marginal damages that appears in the numerator has a minimum for $d = c/aN^3$ equal to

$$-\frac{c^2(2N-1)}{aN^3} + acN,$$

which is positive if

$$\frac{2N-1}{2N^4} < \frac{a^2}{2c}. \tag{46}$$

This inequality holds for $N \geq 2$ if condition $1 < a^2/2c$ is satisfied. Thus, the effective investment is zero for $d = 0$ and positive and increasing for $d \in (0, d_d^s)$. Moreover, $\lim_{d \to d_d^s} y^s(d) = +\infty$. Then, there exists a unique value for $d$ lower than $a/2$ that satisfies that

$$y^s = \frac{N^2d^s(aN-(2N-1)d^s)}{c-(2N-1)N^2(d^s)^2} = 1.$$

Solving this expression for $d$, we obtain $d_N^s = c/aN^3$ in the main text that is just the minimum of the polynomial of second degree in marginal damages that appears in the numerator of (45).

**B) Proof of Proposition 3**

Using (23) and (25) we obtain the difference between the effective investments

$$y^s - y^f = \frac{N^2 d(aN-(2N-1)d)}{c-(2N-1)N^2d^2} - \frac{d(1+\gamma(N-1))(a(1+\gamma(N-1))-d(1+2\gamma(N-1)))}{c-(1+\gamma(N-1))(1+2\gamma(N-1))d^2}$$
$$= \frac{(1-\gamma)(1+\gamma(N-1))N^2ad^2 - b_0(\gamma)cd + b_1(\gamma)ac}{(c-(2N-1)N^2d^2)(c-(1+\gamma(N-1))(1+2\gamma(N-1))d^2)},$$

where

$$b_0(\gamma) = 2N^2+N+1-2(N-1)\gamma^2-3\gamma > 0 \text{ for } \gamma \in (0,1),$$
$$b_1(\gamma) = N^2+N+1-(N-1)\gamma-2\gamma > 0 \text{ for } \gamma \in (0,1).$$

In the numerator we have a polynomial of second degree in $d$ with a polynomial equation that according to the Descartes' rule of signs could have two positive roots. In that case, if the polynomial is positive for $d_N^s$ and its slope is negative at this value, we can conclude that the numerator is positive for all $d$ in the interval $(0, d_N^s)$. Substituting in the numerator $d$ for $d_N^s = c/aN^3$, the following expression is obtained

$$\frac{c}{aN^4}(N^2 - N\gamma^2 + N + (\gamma-1)^2)(a^2N^4 - (2N-1)c),$$

which is positive if

$$\frac{2N-1}{2N^4} < \frac{a^2}{2c}.$$

Observe that this condition is the same condition that guarantees that the effective investment of signatories is increasing, see (46). Thus, the numerator of the difference in investments is positive for $d_N^s$, but to conclude that the difference is positive in the interval $(0, d_N^s)$ we need to check whether the slope of the numerator is negative for $d_N^s$. Notice that the polynomial equation could have two positive roots and $d_N^s$ could be larger than the highest of the two roots given a positive value for the polynomial with negative values for lower marginal damages. The first derivative of the numerator is

$$2(1-\gamma)(1+\gamma(N-1))N^2ad - (2N^2+N+1-2(N-1)\gamma^2-3\gamma)c,$$

which substituting $d$ for $d_N^s$ yields

$$\frac{c}{N}(2\gamma^2 (N-1)^2 + \gamma (5N-4) - \left(N + N^2 + 2N^3 - 2\right)),$$

which is negative for $\gamma \in (0,1)$. On the other hand, according to Lemma 1, $d_N^s < d_0^f$ then for $d \in (0, d_N^s)$, the denominator is also positive because the two factors in the denominator are positive in that interval. Then we can conclude that $y^f < y^s$ for $d \in (0, d_N^s)$. For $[d_N^s, d_0^f)$, $y^f < y^s$ because $y^f$ is lower than the unity and $y^s$ is equal to the unity. Finally, for $d = d_0^f$, both investments are equal to the unity.

**C) The equation (41) has a unique positive solution lower than $d_{N-1}^f$ when $\gamma \geq 0.71$**

As the coefficient of the first term of the polynomial of third degree in the left-hand side of the equation is positive and the independent term is negative, if we show that the polynomial is increasing for all $d$ we can conclude that the equation has only one positive solution. The first derivative with respect to $d$ of the polynomial is

$$3(N-1)^2(1-\gamma^2)(2(N-1)\gamma^2 + \gamma - (N-1))ad^2 - 2(1-\gamma)(2(N-1)^2\gamma - 1)cd$$
$$+((N-1)\gamma^2 + (N^2 - 2N + 2)\gamma + (N-1)^3)ac, \tag{47}$$

where the first coefficient is positive, the second one is negative and the independent term is positive. The minimum value of (47) is given by the following expression

$$-\frac{(1-\gamma)(2(N-1)^2\gamma - 1)^2 c^2}{3(1+\gamma)(N-1)^2(2(N-1)\gamma^2 + \gamma - (N-1))a}$$
$$+((N-1)\gamma^2 + (N^2 - 2N + 2)\gamma + (N-1)^3)ac,$$

which is positive if only if

$$\frac{a^2}{2c}$$
$$> \frac{(1-\gamma)(2(N-1)^2\gamma - 1)^2}{6(1+\gamma)((N-1)\gamma^2 + (N^2 - 2N + 2)\gamma + (N-1)^3)(N-1)^2(2(N-1)\gamma^2 + \gamma - (N-1))}. \tag{48}$$

As we have assumed that $a^2 > 2c$, this condition holds if

$$1 - \frac{(1-\gamma)(2(N-1)^2\gamma - 1)^2}{6(1+\gamma)((N-1)\gamma^2 + (N^2 - 2N + 2)\gamma + (N-1)^3)(N-1)^2(2(N-1)\gamma^2 + \gamma - (N-1))}$$

is positive. Calculating the numerator of this difference, the following polynomial in $\gamma$ is obtained

$$\begin{aligned}
&12\left(N-1\right)^4\gamma^5+6\left(N-1\right)^3\left(2N^2-2N+3\right)\gamma^4\\
&+2\left(N-1\right)^2\left(6N^4-18N^3+20N^2-N-4\right)\gamma^3\\
&+2\left(N-1\right)^2\left(6N^4-24N^3+34N^2-23N+8\right)\gamma^2\\
&-\left(6N^6-36N^5+90N^4-114N^3+68N^2-10N-5\right)\gamma\\
&-(6N^6-36N^5+90N^4-120N^3+90N^2-36N+7),
\end{aligned}$$

where all the polynomials in $N$ are positive for $N\geq 2$. This polynomial in $\gamma$ is negative for $\gamma=0$ and positive for $\gamma=1$ and that the coefficients only change the sign once. For these reasons, the polynomial equation has only one positive root that is lower than the unity. Next, we check the value of the polynomial for $\gamma=0.71$. Substituting above $\gamma$ for 0.71 yields

$$0.08\,N^6+6.\,84N^5-30.\,83N^4+56.\,97N^3-49.\,60N^2-49.\,60N^2+16.\,90N-0.66>0 \text{ for } N\geq 2.$$

This means that the value of $\gamma$ that makes the polynomial zero is lower than 0.71 and therefore the polynomial is positive for $\gamma\in[0.71,1)$. Thus, we can conclude that if $2c<a^2$, the condition (48) is satisfied and that the first derivative (47) is positive for all $d$. In this case, the polynomial in the left-hand side of equation (41) is increasing in $d$ and strictly convex for values of $d$ larger than the value that yields the minimum value for the first derivative.[18] Then the polynomial equation has only one positive root because the independent term is negative. This positive root is $d^s_{N-1}$ which is lower than $d^f_{N-1}$ since as we have shown in the main text, the value of the polynomial in $d$ is positive for $d^f_{N-1}$.

**D) $d^s_{N-1}$ and $d^-_{nx}$ are decreasing with respect to $\gamma$**

[18] Take into account that when the first derivative reaches a minimum, the second derivative is zero and the polynomial on the left-hand side of equation (41) presents an inflexion point. On the left, the polynomial is an increasing concave function of $\gamma$ and on the right it is an increasing convex function of $\gamma$.

The relationship between $d^s_{N-1}$ and $\gamma$ is defined by the implicit function $F(d^s_{N-1}, \gamma) = 0$ given by equation (41). Then, according to the implicit function theorem $\partial d^s_{N-1}/\partial\gamma = -\partial F/\partial\gamma/\partial F/\partial d^s_{N-1}$. The denominator of this quotient is positive because $F$ is increasing in $d^s_{N-1}$ since $d^s_{N-1}$ is the lowest root of equation (41) that has a negative independent term. On the other hand, the numerator is

$$\frac{\partial F}{\partial\gamma} = d((N-1)^2(-8\gamma^3(N-1) - 3\gamma^2 + 6\gamma(N-1) + 1)ad^2$$

$$+(4\gamma(N-1)^2 - (2N^2 - 4N + 3)cd + (2(N-1)\gamma + N^2 - 2N + 2)ac),$$

where

$$-8\gamma^3(N-1) - 3\gamma^2 + 6\gamma(N-1) + 1 > 0 \text{ for } \gamma \leq 0.71.$$

The second term is negative for low values of $\gamma$ and the third term is positive. When, the second term is negative the minimum value of the derivative is given by the following expression

$$-\frac{(4\gamma(N-1)^2 - (2N^2 - 4N + 3)^2c^2}{4(N-1)^2(-8\gamma^3(N-1) - 3\gamma^2 + 6\gamma(N-1) + 1)a}$$
$$+(2(N-1)\gamma + N^2 - 2N + 2)ac,$$

which is positive provided that

$$\frac{a^2}{2c} > \frac{(4\gamma(N-1)^2 - (2N^2 - 4N + 3)^2}{8(N-1)^2(-8\gamma^3(N-1) - 3\gamma^2 + 6\gamma(N-1) + 1)(2(N-1)\gamma + N^2 - 2N + 2)} > 0.$$

As we have assumed that $a^2 > 2c$, this condition is satisfied if

$$1 - \frac{(4\gamma(N-1)^2 - (2N^2 - 4N + 3)^2}{8(N-1)^2(-8\gamma^3(N-1) - 3\gamma^2 + 6\gamma(N-1) + 1)(2(N-1)\gamma + N^2 - 2N + 2)}$$

is positive.

The numerator of this difference is

$$-128(N-1)^4\gamma^4 - 16(N-1)^3(4N^2 - 8N + 11)\gamma^3 + 8(N-1)^2(7N^2 - 14N + 4)\gamma^2$$

$$+8(N-1)^2(6N^3 - 16N^2 + 22N - 11)\gamma + 4N^4 - 16N^3 + 28N^2 - 24N + 7, \tag{49}$$

where all the polynomials in $N$ are positive for $N \geq 2$. Notice that this polynomial in $\gamma$ is positive for $\gamma = 0$ and negative for $\gamma = 1$ and that the coefficients only change the sign once. Then the polynomial equation has only one positive root that is lower than the unity. Next, we check the value of polynomial for $\gamma = 0.71$. Substituting above $\gamma$ for 0.71 yields

$$11.17N^5 - 44.81N^4 + 72.84N^3 - 63.84N^2 + 32.52N - 8.88 > 0 \text{ for } N \geq 2.$$

The sign of this inequality establishes that the value of $\gamma$, which makes the polynomial (49) zero, is larger than 0.71 and that therefore the polynomial is positive for all $\gamma \in (0, 0.71)$ and consequently for all $\gamma \in (0, \gamma_1(N))$, since as has been defined in the main text $\gamma_1(N) < 0.71$. Thus, we can conclude that the minimum value of $\partial F/\partial\gamma$ is positive for $\gamma \in (0, \gamma_1(N))$ so that $\partial F/\partial\gamma$ is positive in that interval and by the implicit function theorem we obtain that $\partial d^s_{N-1}/\partial\gamma$ is negative.

On the other hand, the relationship between $d^-_{nx}$ and $\gamma$ is defined in the main text by the explicit function (44). The first derivative of this function is

$$\frac{\partial d^-_{nx}}{\partial\gamma} = \frac{\left(aN(N-1) - \frac{a^2N^2(N-1)(1+\gamma(N-1))+2c(N-2\gamma)}{(N^2(1+\gamma(N-1))^2a^2-4(1-\gamma)(N-1-\gamma)c)^{1/2}}\right)2(1-\gamma)(N-1-\gamma)}{4(1-\gamma)^2(N-1-\gamma)^2}$$

$$+\frac{(N(1+\gamma(N-1))a - (N^2(1+\gamma(N-1))^2a^2 - 4(1-\gamma)(N-1-\gamma)c)^{1/2})2(N-2\gamma)}{4(1-\gamma)^2(N-1-\gamma)^2}.$$

Let us assume that the numerator is positive or zero. Rearranging and collecting terms in the numerator this assumption implies the following inequality

$$aN(N^2 - N + 1 - (N-1)\gamma^2 - 2\gamma)(N^2(1+\gamma(N-1))^2a^2 - 4(1-\gamma)(N-1-\gamma)c)^{1/2}$$

$$\geq a^2N^2(1+\gamma(N-1))(2(N-1)(1-\gamma)(N-1-\gamma) + (1+\gamma(N-1))(N-2\gamma))$$

$$-2c(1-\gamma)(N-1-\gamma)(N-2\gamma),$$

where the right-hand side is positive if $2c < a^2$.[19] Squaring and rearranging terms gives the following contradiction

$$0 \geq a^4N^4(N-1)(1+\gamma(N-1))^2(\gamma^3(N-1) + \gamma^2(N+3) - \gamma\left(N + N^2 + N^3 - 1\right)$$

$$+ (N-1)\left(-4N + 3N^2 + 3\right)) + 4a^2c(1-\gamma)^2N^2(N-1)^2(N-1-\gamma)(\gamma - N + 1)^2$$

[19] We omit the details because the same procedure that that used above to show the minimum value of the $\partial F/\partial\gamma$ is positive is applied.

$$+4(1-\gamma)(N-1-\gamma)^2(N-2\gamma)^2c^2, \tag{50}$$

where

$$\gamma^3(N-1)+\gamma^2(N+3)-\gamma\left(N^3+N^2+N-1\right)+(N-1)\left(3N^2-4N+3\right)>0 \tag{51}$$

for $\gamma \in (0,1)$. Observe that the first derivative of this polynomial in $\gamma$ has only one positive root and that the second derivative is positive for all $\gamma > 0$. Thus, (51) is strictly convex for $\gamma > 0$ and has a minimum for the positive root of the first derivative. Moreover, (51) is $2N(N-2)^2$ for $\gamma = 1$ and the first derivative is $-(N^3+N^2-4N-4)$, that is negative for $N > 2$. Then, if the polynomial equation of (51) has two positive roots, the lowest root must be larger than the unity and therefore (51) is positive in the interval $(0,1)$. The contradiction allows us to refuse the assumption that $\partial d^-_{nx}/\partial\gamma \geq 0$ and consequently to conclude that $d^-_{nx}$ decreases with $\gamma$.

Finally, we show that $d^-_{nx}$ takes a positive finite value for $\gamma = 1$. Substituting $\gamma$ for 1 in (44) the indeterminate form 0/0 is obtained. This indeterminate form can be solved applying the L'Hôpital's rule

$$\lim_{\gamma\to 1} d^-_{nx} = \lim_{\gamma\to 1} \frac{aN(N-1) - \frac{a^2N^2(N-1)(1+\gamma(N-1))+2c(N-2\gamma)}{(N^2(1+\gamma(N-1))^2a^2-4(1-\gamma)(N-1-\gamma)c)^{1/2}}}{-2(N-2\gamma)} = \frac{c}{aN^2} > 0.$$

# References


[1] Barrett, S., 1994. Self-Enforcing International Environmental Agreements. Oxford Economic Papers 46, 878-894.

[2] Barrett, S., 2006. Climate Treaties and "Breakthrough" Technologies. American Economic Review 96, 22-25.

[3] Barrett, S., 2009. The Coming Global Climate-Technology Revolution. Journal of Economic Perspectives 23, 53-75.

[4] Battaglini, M., Harstad, B., 2015. Participation and Duration of Environmental Agreements. Journal of Political Economy, forthcoming.

[5] Beccherle, J., Tirole, J., 2011. Regional Initiatives and the Cost of Delaying Binding Climate Change Agreements. Journal of Public Economics 95, 1339-1348.

[6] Benchekroun, H. Chaudhuri, A.R. 2014. Cleaner Technologies and the Stability of International Environmental Agreements. Journal of Public Economic Theory, forthcoming.

[7] Buchholz, W., Konrad, K.A., 1994. Global Environmental Problems and the Strategic Choice of Technology. Journal of Economics 60, 299-321.

[8] Buchner, B., Carraro, C., 2004. Economic and Environmental Effectiveness of a Technology-Based Climate Protocol. Climate Policy 4, 229-248.

[9] Carraro, C., Siniscalco, D. 1993. Strategies for the International Protection of the Environment. Journal of Public Economics 52, 309-328.

[10] Carraro, C., Siniscalco, D., 1997. R&D Cooperation and the Stability of International Environmental Agreements. In International Environmental Negotiations: Strategic Policy Issues, edited by C. Carraro. Edward Elgar Publishing.

[11] Chander, P., Tulkens, H., 1995. A Core-Theoretic Solution for the Design of Cooperative Agreements on Transfrontier Pollution. International Tax and Public Finance 2, 279-294.

[12] Cornes, R., Sandler, T., 1996. The Theory of Externalities, Public Goods and Club Goods, 2nd ed. Cambridge University Press.

[13] d'Aspremont, C., Jacqemin, J., Gadszeweiz, J., Weymark, J.A., 1983. On the Stability of Collusive Price Leadership. Canadian Journal of Economics 16, 17- 25.

[14] de Coninck, H., Fischer, C., Newell, R.G., Ueno, T., 2008. International Technology-Oriented Agreements to Address Climate Change. Energy Policy 36, 335-356.

[15] El-Sayed, A., Rubio, S.J., 2014. Sharing R&D Investments in Cleaner Technologies to Mitigate Climate Change. Resource and Energy Economics 38, 168-180.

[16] Goeschl, T., Perino, G., 2015. The Climate Policy Hold-Up: Green Technologies, Intellectual Property Rights, and the Abatement Incentives of International Agreements. Discussion Paper Series no. 591, Department of Economics, University of Heidelberg.

[17] Harstad, B., 2012. Climate Contracts: A Game of Emissions, Investments, Negotiations, and Renegotiations. Review of Economic Studies 79, 1527-1557.

[18] Harstad, B., 2015. The Dynamics of Climate Agreements. Journal of the European Economic Association, forthcoming.

[19] Helm, C., Schmidt, R.C., 2015. Climate Cooperation with Technology Investments and Border Carbon Adjustment. European Economic Review 75, 112-130.

[20] Hoel, M., de Zeeuw, A., 2010. Can a Focus on Breakthrough Technologies Improve the Performance of International Environmental Agreements? Environmental and Resource Economics 47, 395-406.

[21] Hoel, M., de Zeeuw, A., 2014. Technology Agreements with Heterogeneous Countries. In Toward a New Climate Agreement. Conflict, Resolution and Governance, edited by T. L. Cherry, J. Hovi and D. M. McEvoy. Routledge.

[22] Hong, F., Karp, L., 2012. International Environmental Agreements with Mixed Strategies and Investment. Journal of Public Economics 96, 685-697.

[23] Kamien, M.I., Muller, E., Zang, I., 1992. Research Joint Ventures and R&D Cartels. American Economic Review 82, 1293-1306.

[24] Kemfert, C., 2004. Climate Coalitions and International Trade: Assessment of Cooperation Incentives by Issue Linkage. Energy Policy 32, 455-465.

[25] Lessmann, K., Edenhofer, O., 2011. Research Cooperation and International Standards in a Model of Coalition Stability. Resource and Energy Economics 33, 36-54.

[26] Nagashima, M., Dellink, R., 2008. Technology Spillovers and Stability of International Climate Coalitions. International Environmental Agreements: Politics, Law and Economics 8, 343-365.

[27] Nagashima, M., Weikard, H-P., de Bruin, K., Dellink, R., 2011. International Climate Agreements under Induced Technological Change. Metroeconomica 62, 612-634.

[28] Nordhaus, W., 2015. Climate Clubs: Overcoming Free-Riding in International Climate Policy. American Economic Review 105, 1339-1370.

[29] Ruis, A., de Zeeuw, A., 2010. International Cooperation to Combat Climate Change. Public Finance and Management 10, 379-404.

[30] Schmidt, R.C., Strausz, R., 2014. On the Timing of Climate Agreements. Environmental and Resource Economics, forthcoming.

[31] Ulph, A., Ulph, D. 2007. Climate Change - Environmental and Technology Policies in a Strategic Context. Environmental and Resource Economics 37, 159-180.

[32] Urpelainen, J., 2014. Sinking Costs to Increase Participation: Technology Deployment Agreements Enhance Climate Cooperation. Environmental Economics and Policy Studies 16, 229-240.

[33] van der Pol, T., Weikard, H-P., van Ireland, E., 2012. Can Altruism Stabilise International Climate Agreements? Ecological Economics 81, 112-120.

| $n$ | $x^s$ | $x^f$ | $y^s$ | $y^f$ | $q^s$ | $q^f$ |
|---|---|---|---|---|---|---|
| 1 | | $1.44962 \times 10^{-3}$ | | 0.0021 | | 9.9980042 |
| 2 | 0.004 | $1.44966 \times 10^{-3}$ | 0.010 | 0.0024 | 9.99802 | 9.9980048 |
| 3 | 0.010 | $1.44968 \times 10^{-3}$ | 0.031 | 0.0034 | 9.99806 | 9.9980067 |
| 4 | 0.017 | $1.44970 \times 10^{-3}$ | 0.069 | 0.0052 | 9.99813 | 9.9980105 |
| 5 | 0.026 | $1.44972 \times 10^{-3}$ | 0.131 | 0.0083 | 9.99826 | 9.9980166 |
| ► 6 | 0.037 | $1.44974 \times 10^{-3}$ | 0.223 | 0.0128 | 9.99844 | 9.9980256 |
| 7 | 0.050 | $1.44976 \times 10^{-3}$ | 0.350 | 0.0191 | 9.99870 | 9.9980382 |
| 8 | 0.065 | $1.44978 \times 10^{-3}$ | 0.518 | 0.0274 | 9.99903 | 9.9980549 |
| 9 | 0.081 | $1.44980 \times 10^{-3}$ | 0.733 | 0.0381 | 9.99946 | 9.9980762 |
| 10 | 0.100 | | 1.000 | | 10.0000 | |

TABLE 1A. The stable coalition for $d = 0.002$

| $n$ | $e^s$ | $e^f$ | $E$ | $W^s$ | $W^f$ | $W^f - W^s$ |
|---|---|---|---|---|---|---|
| 1 | | 9.976 | 99.76 | | 49.8004 | |
| 2 | 9.896 | 9.973 | 99.58 | 49.8006 | 49.8008 | 0.0002 |
| 3 | 9.692 | 9.964 | 98.82 | 49.8013 | 49.8023 | 0.0010 |
| 4 | 9.306 | 9.945 | 96.89 | 49.8032 | 49.8061 | 0.0029 |
| 5 | 8.683 | 9.915 | 92.99 | 49.8071 | 49.8139 | 0.0068 |
| ► 6 | 7.764 | 9.869 | 86.07 | 49.8140 | 49.8278 | 0.0138 |
| 7 | 6.494 | 9.806 | 74.88 | 49.8251 | 49.8502 | 0.0251 |
| 8 | 4.815 | 9.723 | 57.97 | 49.8420 | 49.8840 | 0.0420 |
| 9 | 2.669 | 9.617 | 33.64 | 49.8663 | 49.9326 | 0.0663 |
| 10 | 0.000 | | 0.00 | 49.9000 | | |

TABLE 1B. The stable coalition for $d = 0.002$

| $n$ | $x^s$ | $x^f$ | $y^s$ | $y^f$ | $q^s$ | $q^f$ |
|---|---|---|---|---|---|---|
| 1 | | 0.07660 | | 0.111 | | 9.9049 |
| 2 | 0.254 | 0.07667 | 0.539 | 0.129 | 9.9507 | 9.9068 |
| 3 | 0.324 | 0.07679 | 1 | 0.148 | 10 | 9.9089 |
| 4 | 0.244 | 0.07684 | 1 | 0.144 | 10 | 9.9085 |
| 5 | 0.196 | 0.07688 | 1 | 0.141 | 10 | 9.9081 |
| 6 | 0.164 | 0.07693 | 1 | 0.137 | 10 | 9.9077 |
| 7 | 0.141 | 0.07698 | 1 | 0.134 | 10 | 9.9073 |
| 8 | 0.124 | 0.07702 | 1 | 0.130 | 10 | 9.9069 |
| 9 | 0.117 | 0.07707 | 1 | 0.126 | 10 | 9.9065 |
| ▶ 10 | 0.10 | | 1 | | 10 | |

TABLE 2A. The stable coalition for $d$=0.107

| $n$ | $e^s$ | $e^f$ | $E$ | $W^s$ | $W^f$ | $W^f - W^s$ |
|---|---|---|---|---|---|---|
| 1 | | 8.80 | 88.05 | | 40.51 | |
| 2 | 4.58 | 8.63 | 78.20 | 40.98 | 41.57 | 0.59 |
| 3 | 0 | 8.44 | 59.07 | 42.63 | 43.62 | 0.99 |
| 4 | 0 | 8.47 | 50.84 | 43.96 | 44.50 | 0.54 |
| 5 | 0 | 8.51 | 42.54 | 45.06 | 45.38 | 0.32 |
| 6 | 0 | 8.54 | 34.17 | 46.07 | 46.28 | 0.21 |
| 7 | 0 | 8.58 | 25.74 | 47.05 | 47.18 | 0.13 |
| 8 | 0 | 8.61 | 17.22 | 48.00 | 48.09 | 0.09 |
| 9 | 0 | 8.65 | 8.65 | 48.95 | 49.01 | 0.06 |
| ▶ 10 | 0 | | 0 | 49.90 | | |

TABLE 2B. The stable coalition for $d$=0.107.

| $n$ | $x^s$ | $x^f$ | $y^s$ | $y^f$ | $q^s$ | $q^f$ |
|---|---|---|---|---|---|---|
| 1 | | 0.1768 | | 0.2564 | | 9.8141 |
| 2 | 0.4645 | 0.1774 | 1 | 0.2860 | 10 | 9.8215 |
| 3 | 0.3126 | 0.1776 | 1 | 0.2778 | 10 | 9.8194 |
| 4 | 0.2366 | 0.1778 | 1 | 0.2696 | 10 | 9.8174 |
| 5 | 0.1911 | 0.1780 | 1 | 0.2614 | 10 | 9.8153 |
| 6 | 0.1607 | 0.1782 | 1 | 0.2532 | 10 | 9.8133 |
| 7 | 0.1390 | 0.1784 | 1 | 0.2449 | 10 | 9.8112 |
| 8 | 0.1228 | 0.1786 | 1 | 0.2367 | 10 | 9.8092 |
| 9 | 0.1101 | 0.1788 | 1 | 0.2284 | 10 | 9.8071 |
| ▶ 10 | 0.1000 | | 1 | | 10 | |

TABLE 3A. The stable coalition for $d$=0.25

| $n$ | $e^s$ | $e^f$ | $E$ | $W^s$ | $W^f$ | $W^f - W^s$ |
|---|---|---|---|---|---|---|
| 1 | | 7.30 | 73.00 | | 31.42 | |
| 2 | 0 | 7.01 | 56.10 | 33.81 | 35.64 | 1.83 |
| 3 | 0 | 7.09 | 49.64 | 36.61 | 37.26 | 0.65 |
| 4 | 0 | 7.17 | 43.02 | 38.68 | 38.91 | 0.23 |
| 5 | 0 | 7.25 | 36.25 | 40.57 | 40.60 | 0.03 |
| 6 | 0 | 7.33 | 29.31 | 42.41 | 42.34 | −0.07 |
| 7 | 0 | 7.41 | 22.22 | 44.25 | 44.11 | −0.14 |
| 8 | 0 | 7.49 | 14.97 | 46.12 | 45.92 | −0.20 |
| 9 | 0 | 7.57 | 7.57 | 47.99 | 47.77 | −0.22 |
| ▶ 10 | 0 | | 0 | 49.90 | | |

TABLE 3B. The stable coalition for $d$=0.25.

**NOTE DI LAVORO DELLA FONDAZIONE ENI ENRICO MATTEI**

**Fondazione Eni Enrico Mattei Working Paper Series**

**Our Note di Lavoro are available on the Internet at the following addresses:**
http://www.feem.it/getpage.aspx?id=73&sez=Publications&padre=20&tab=1
http://papers.ssrn.com/sol3/JELJOUR_Results.cfm?form_name=journalbrowse&journal_id=266659
http://ideas.repec.org/s/fem/femwpa.html
http://www.econis.eu/LNG=EN/FAM?PPN=505954494
http://ageconsearch.umn.edu/handle/35978
http://www.bepress.com/feem/
http://labs.jstor.org/sustainability/

**NOTE DI LAVORO PUBLISHED IN 2016**

| | | |
|---|---|---|
| ET | 1.2016 | Maria Berrittella, Carmelo Provenzano: An Empirical Analysis of the Public Spending Decomposition on Organized Crime |
| MITP | 2.2016 | Santiago J. Rubio: Sharing R&D Investments in Breakthrough Technologies to Control Climate Change |