

Climate Treaties and Breakthrough Technologies in a Dynamic Framework¹

1st June 2010

Alejandro Caparrós

Consejo Superior de Investigaciones Científicas (CSIC). Institute for Public Goods and Policies (IPP). Albasanz 26, 28037 Madrid, Spain. E-mail: alejandro.caparros@cchs.csic.es

Raúl López-Castro

Universidad Complutense Madrid. Dpto. de Fundamentos del Análisis Económico I.

Campus de Somosaguas. 28223, Pozuelo de Alarcón, Madrid, Spain. E-mail:

raul.lc@wanadoo.es

¹Preliminary version, please do not cite or quote.

Climate Treaties and Breakthrough Technologies in a Dynamic Framework²

1st June 2010

Abstract: We analyze a dynamic monotone game with discounting over the implementation of a pure public good. After discussing the general model, we apply the model to the analysis of a breakthrough technology with increasing returns. Our results suggest that if the number of countries (agents) is small, the technology will be adopted without the need of any treaty. Nevertheless, the technology will not necessarily be adopted immediately and the potential delay is proportional to the number of countries. On the contrary, if the number of countries is large the technology may never be adopted. In this case, a treaty can improve the situation since a small stable coalition of countries would be interested in adopting the technology and launching the process that ultimately leads to the universal adoption of the technology.

JEL Classification: C73, Q54, F59.

Key words: monotone games, breakthrough technologies, international environmental agreements, game theory, dynamic coordination games.

²Preliminary version, please do not cite or quote.

1 Introduction

Game theoretic analysis of international cooperation on climate change has generally reached rather negative conclusions, showing that only a handful of countries would join a treaty focused on abatement targets (Carraro and Siniscalco, 1993; Barrett, 1994). Recent research has raised the question whether or not focusing on R&D and “breakthrough” technologies may change this conclusion. Barrett (2006) has shown that the answer is yes if this breakthrough technologies exhibit increasing returns to scale, while the answer is no otherwise. Hoel and Zeeuw (2009), by introducing the assumption that the adoption costs of a breakthrough technology vary with the level of R&D, find that even without increasing returns to scale a focus on R&D may yield better results than those obtained focusing on abatement targets (the path followed by the Kyoto Protocol).

The papers just mentioned are static games where all countries decide simultaneously whether or not to invest in R&D. More precisely, these papers assume two or three stages (the precise stages vary from one paper to another) but adoption decision are taken simultaneously. For example, in Barrett (2006) first countries decide whether to be party to a technology adoption agreement, then signatories decide collectively whether to adopt the new technology and finally non-parties decide individually whether to adopt the technology. However, adoption of breakthrough technologies and investment in R&D decisions will, in reality, not follow such a simultaneous pattern. Windmills, or electric cars, or second generation biofuels are going to have different cost in different moments in time, at least partly due to the efforts done by those who decided to move on earlier.

To analyze this issue we propose a model that represents the adoption of breakthrough

technologies with increasing returns to scale (the most favorable case discussed above) in a dynamic context. We first present a general model and then discuss the implications in our framework of the X-type technology proposed in Barrett (2006). Our model builds on Gale (1995) by taking into account the pure public good nature of the problem under consideration (Gale (2001) analyzes environments with positive spillovers but without discounting). In our framework, countries do not only benefit once they have decided to invest since abatement efforts done by others are benefitting them from the moment when they were done. On the costs side, maintenance costs are independent of previous investments but installation costs depend on the number of other countries which have adopted the new technology.

2 The model

2.1 Implementation phase

There are N identical players ($i = 1, 2, \dots, N$) which have to decide when to start investing in the new technology. Once they decide to invest, all countries benefit from the abatement obtained but only those which have invested incur in costs (initial costs at the year when the investment starts and maintenance costs thereafter). Once a country has decided to adopt the new technology it will continue to use it indefinitely (i.e. the initial costs are large enough to preclude starting the investment and then abandoning the technology). This implies to have a monotone game (Gale, 1995). Formally, the play of the game occurs at a countable set of dates, indexed by $t = 1, \dots, \infty$, and player i 's action at any moment in time is defined by $x_{i,t} = \{0, 1\}$, with $x_{i,t} \geq x_{i,t-1} \forall t \geq 1$ and $x_{i,0} = 0$. The discount factor is δ , with

$0 < \delta < 1$, and the proportion of countries that have already invested in t is defined by:

$$\alpha_t = \frac{\sum_{i=1}^N x_{it}}{N} \quad (1)$$

$B(\alpha)$ is the benefit function, increasing in the proportion of players that have already invested. All players benefit as soon as $\alpha_t > 0$, not only those investing in the new technology. Costs $C(\alpha)$ are assumed to be a decreasing function of the countries that have already invested. To simplify we assume that there are only costs in the initial period of investment s (constant maintenance costs are assumed to be capitalized in this value). Thus, country i 's net benefit function is:

$$\Pi_{i,1} = \sum_{t=1}^{\infty} \delta^{t-1} B(\alpha_t) - C(\alpha_s) \quad (2)$$

We make the following assumptions:

A.I: $B(0) = 0$ and $C(0) = 0$

A.II: $B(1) - B\left(1 - \frac{1}{N}\right) > C(1)(1 - \delta)$

A.III: $B\left(\frac{1}{N}\right) < C\left(\frac{1}{N}\right)(1 - \delta)$

A.IV: $B(\alpha)$ is continuous, increasing and non-negative.

A.V: $C(\alpha)$ is continuous, decreasing and non-negative.

Assumption I implies that if nobody invests nobody gains, and that there are no economies of scale if nobody invested previously. Assumptions II and III just ensure that there is an economically interesting problem. A.II implies that when there is only one player remaining he will be interested in investing. A.III means that for a singleton the abatement achieved would not cover the costs (i.e. there is no incentive to invest unilaterally). The last two

assumptions just allow the mathematical treatment of the model and are satisfied by most benefit or cost functions proposed in previous literature.

These assumptions and (2) ensure that for some $0 < \alpha^* < 1$,

$$\frac{B_t(\alpha^*) - B_t\left(\alpha^* - \frac{1}{N}\right)}{(1 - \delta)} = C(\alpha^*) \quad (3)$$

Thus, α^* represents the proportion of countries that need to invest before it becomes profitable doing so.

We have assumed that the decisions are taken at countable dates, i.e., countries need some time to take their decision. The next Theorem shows that the amount of time needed is in fact relevant since it has a strong impact on the delay until the agreement is reached (the delay is obviously costly since we have a positive discount rate).

Theorem 1 *As the period length tends to 0, the delay until all players invest and hence reach the full investment equilibrium tends to 0.*

Proof. Let $\tau > 0$ denote the period length. The revenue flow per period is $B(\alpha)\tau$ and the discount factor $\delta = e^{-\rho\tau}$, with $\rho > 0$ being the fixed discount rate. Let Γ_n denote the equivalence class of subgames in which exactly n players have already invested and let $T_n(\tau)$ denote the supremum of the time taken for all players to invest, taken over all SPE of $\Gamma_n(\tau)$. As induction hypothesis, assume that for $n = k, \dots, N$, $T_n(\tau) \rightarrow 0$ as $\tau \rightarrow 0$. This is clearly true for $k = N - 1$ since, by assumption II, once that all but one players have invested it is a dominant strategy for the remaining player to invest as soon as he can (in other words, he would wait at most one period). Thus: $T_{N-1}(\tau) = \tau$. Suppose that the

induction hypothesis is true for some arbitrary $k < N$ (at least the one previously shown). Then we need to show that it also holds for $k - 1$. If some players invest in the game Γ_{k-1} , then they will precipitate a subgame Γ_n with $n > k - 1$. And by the induction hypothesis, delay in this game Γ_n is bounded by $T_n(\tau)$, so the induction hypothesis can be proved for $k - 1$ by establishing a bound on the time until the first player invests in the subgame Γ_{k-1} (and then precipitates a subgame Γ_k that is bounded as shown before). This can be done by contradiction. For any SPE of Γ_{k-1} , let d denote the time taken for the first player to invest (that is, the delay). We need to prove that as τ converges to zero, d converges uniformly to zero for all SPE. Suppose the contrary: then we can find a sequence of period lengths $\{\tau_r\}$ and corresponding equilibria $\{f_r\}$ with the property that τ_r converges to zero as $r \rightarrow \infty$ and $d_r \geq d > 0$ for all r sufficiently large (that is, there exists some equilibria when delay and period length tends to zero). Consider some fixed but arbitrary equilibrium f_r and the following deviation from the equilibrium strategies: some player invests at the first date, thus precipitating the subgame Γ_k beginning at the second date. By hypothesis, once Γ_k has started, all the players will invest within $T_k(\tau_r)$ periods. The worst situation for him is that all the others do not invest until $T_k(\tau_r)$ periods have passed. Thus, in the first $T_k(\tau_r)$ periods he will receive at least $\sum_{t=1}^{T_k(\tau_r)} e^{-\rho t \tau_r} B_t(\alpha_t) \tau_r - C \left(\frac{k}{N} \right) \tau_r$, with α_t determined by how many countries have invested until the date, that is $\frac{k}{N} \leq \alpha_t \leq 1$. Once all the others invest and the fully investment equilibrium is reached, he would obtain $\sum_{t=T_k(\tau_r)+\tau_r}^{\infty} e^{-\rho t \tau_r} B(1) \tau_r$. Therefore, a deviating player would obtain at least:

$$e^{-\rho[T_k(\tau_r)+\tau_r]} \frac{B(1)\tau_r}{1 - e^{-\rho\tau_r}} + \sum_{t=1}^{T_k(\tau_r)} e^{-\rho t \tau_r} [B_t(\alpha_t)] \tau_r - C \left(\frac{k}{N} \right) \tau_r \quad (4)$$

On the other hand, in the original equilibrium the best that can be hoped for is that all players invest after d_r time units. In that situation, the player does not invest until d_r units of time have passed and his payoff cannot exceed $\sum_{t=1}^{d_r} e^{-\rho\tau_r} B\left(\frac{k-1}{N}\right) \tau_r$ until d_r time units pass and $\sum_{d_r}^{\infty} e^{-\rho\tau_r} B(1) \tau_r - e^{-\rho\tau_r} C(1) \tau_r$ thereafter. Hence, the total payoff cannot be more than:

$$\sum_{t=1}^{d_r} e^{-\rho\tau_r} B\left(\frac{k-1}{N}\right) \tau_r + e^{-\rho d_r} \frac{B(1) \tau_r}{1 - e^{-\rho\tau_r}} - e^{-\rho\tau_r} C(1) \tau_r \quad (5)$$

To prevent a profitable deviation, the payoff from deviating cannot be higher than the payoff of the original equilibrium, that is (4) \leq (5). Taking limits as $r \rightarrow \infty$ yields (note that as $r \rightarrow \infty$, $\tau_r \rightarrow 0$):

$$\frac{B(1)}{\rho} \leq e^{-\rho d} \frac{B(1)}{\rho}$$

which is the contradiction that we were seeking. Then, a bound for the time taken for a first player to invest in the game Γ_{k-1} has been found and hence, by induction, this proves that $T_n(\tau) \rightarrow 0$ as $\tau \rightarrow 0$, for $n = 0, 1, \dots, N$. ■

The next theorem shows that the number of players involved is also relevant, and that the larger the number of countries involved, the larger the delay in reaching an agreement may be.

Theorem 2 *There exists a SPE in which full investment is reached with a delay proportional to N .*

Proof. Let n^* be the smallest integer greater than $\alpha^* N$ (the number of players that should invest before it is individually profitable doing so) and let Γ_n denote the equivalence class of

subgames in which exactly n players have already invested when the subgame begins, that is, when $t = 1$.

Define the following equilibrium strategies:

- if $n = 1, \dots, n^* - 1$, then the remaining players invest at date $n^* - n$, if there is no deviation.

- if $n \geq n^*$, then the remaining players invest as soon as they can, that is, at date 1.

If these strategies constitute a SPE, then we would have an equilibrium with a delay proportional to N , because n^* is determined by N . We will show that these strategies define a SPE by induction on the number of players. For $n \geq n^* - 1$, investing immediately is a dominant strategy because of the definition of n^* , so it is a SPE for this case. As induction hypothesis, suppose that for every $n = k, \dots, n^* - 1$ these strategies define a SPE of Γ_n . Then, we want to prove that the same holds for $n = k - 1$.

Using the strategies defined above, in Γ_{k-1} every player should invest at date $n^* - k + 1$ if nobody invest earlier, because in such a situation the strategy for the appropriate new subgame would apply. If a player considers deviating, then he will not want to invest later than $n^* - k + 1$, because he would be worse off, so he would only consider deviating before the date determined in the strategies for investing. If he in fact invests before $n^* - k + 1$, he would precipitate the subgame Γ_k , with the other players investing $n^* - k$ periods later. If deviating is profitable, he would like to do it as soon as possible (i.e., in the first period). Then, the subgame Γ_k would start in period 2 of the original game and the remaining players would invest at date $n^* - k + 1$, counted from the first period of the original game. His payoff from deviating would be for the first $n^* - k$ periods $\sum_{t=1}^{n^*-k} \delta^{t-1} B\left(\frac{k}{N}\right) - C\left(\frac{k}{N}\right)$ and $\sum_{t=n^*-k+1}^{\infty} \delta^{t-1} B(1)$ for the remaining periods (where he would obtain the fully cooperative

payoff). Hence, the total actual payoff from deviating is:

$$\sum_{t=1}^{n^*-k} \delta^{t-1} B\left(\frac{k}{N}\right) - C\left(\frac{k}{N}\right) + \sum_{t=n^*-k+1}^{\infty} \delta^{t-1} B(1) \quad (6)$$

On the other hand, if the player does not deviate, he would receive $B\left(\frac{k-1}{N}\right)$ per period until $n^* - k + 1$, with no costs, and $B(1)$ per period thereafter with a cost of $C(1)$ at the time of investment. Thus, the total payment for not deviating is:

$$\sum_{t=1}^{n^*-k} \delta^{t-1} B\left(\frac{k-1}{N}\right) + \sum_{t=n^*-k+1}^{\infty} \delta^{t-1} B(1) - \delta^{n^*-k} C(1) \quad (7)$$

If the strategies define a SPE, we should have that (6) \leq (7). Simplifying and rearranging terms, we need:

$$C\left(\frac{k}{N}\right) - \delta^{n^*-k} C(1) \geq \frac{1 - \delta^{n^*-k}}{1 - \delta} \left(B\left(\frac{k}{N}\right) - B\left(\frac{k-1}{N}\right) \right)$$

Since $\frac{k}{N} < \alpha^*$ we have that $C\left(\frac{k}{N}\right) > \frac{1}{1-\delta} (B\left(\frac{k}{N}\right) - B\left(\frac{k-1}{N}\right))$. Then, we clearly have that $C\left(\frac{k}{N}\right) (1 - \delta) > (1 - \delta^{n^*-k}) (B\left(\frac{k}{N}\right) - B\left(\frac{k-1}{N}\right))$ and also that $C\left(\frac{k}{N}\right) - \delta^{n^*-k} C(1) > C\left(\frac{k}{N}\right) (1 - \delta)$. So, we have shown that this inequality holds with our assumptions and thus, by induction, we have proved that the proposed strategies constitute an equilibrium in every subgame Γ_n . ■

The previous theorem has shown that, as in Gale (1995) for private goods, there exists a SPE in which all players invest at date n^* . As already discussed in Gale (1995), a similar proof can be used to show that there exists a SPE in which all countries invest at date $t = 1, \dots, n^* - 1$. Furthermore, there are also SPEs in which countries invest at different

dates. For example, all players invest at date i , for all $i = 1, \dots, n^* - 2$ and the remaining players invest at date $n^* - 1$. However, although the build up can be slow, given that immediate investment becomes a dominant strategy once cumulative investment reaches n^* , the investment process always ends with a "bang".

Our next result shows that $n^* - n$ is in fact the maximum delay that is always robust, independently of the parameters of the model. Larger delays are possible, but not independently of the parameters of the model, as the next propositions shows.

Proposition 3 *For $n \geq n^* - 1$, any strategy involving delay is not a SPE. For any $n = n^* - x$, with $n^* \geq x > 1$, a strategy to invest in $t = n^* - n + d$, with $d > 1$, is not a SPE of the $\Gamma_{n^* - x}$ subgame if the following condition holds for any $n \geq n^* - x$.*

$$C\left(\frac{n^* - x + 1}{N}\right) - C(1)\delta^{x+d-1} < \sum_1^{x-1} \delta^{t-1} B\left(\frac{n^* - x + 1}{N}\right) + \sum_x^{x-1+d} \delta^{t-1} B(1) - \sum_1^{x-1+d} \delta^{t-1} B\left(\frac{n^* - x}{N}\right) \quad (8)$$

Proof. Investing immediately is a dominant strategy for $n \geq n^* - 1$ because of the definition of n^* . For $x = 2$, deviating implies to obtain $B\left(\frac{n^* - 1}{N}\right)$ in the first period, with a cost of $C\left(\frac{n^* - 1}{N}\right)$, and to obtain full investment in the second period since it will launch the $\Gamma_{n^* - 1}$ subgame where all players invest in one period. On the other hand, following the proposed strategy yields $\sum_1^{1+d} \delta^{t-1} B\left(\frac{n^* - 2}{N}\right) + \sum_{2+d}^{\infty} \delta^{t-1} B(1) - C(1)\delta^{1+d}$. Deviating is profitable if

$$C\left(\frac{n^* - 1}{N}\right) - C(1)\delta^{1+d} < B\left(\frac{n^* - 1}{N}\right) + \sum_2^{1+d} \delta^{t-1} B(1) - \sum_1^{1+d} B\left(\frac{n^* - 2}{N}\right) \quad (9)$$

Thus, if this condition holds the maximum delay in the $\Gamma_{n^* - 2}$ subgame is $n^* - (n^* - 2) = 2$. In the $\Gamma_{n^* - 3}$ subgame, and assuming that condition (9) holds, deviating implies

to receive $B(\frac{n^*-2}{N})$ in the first period, with a cost of $C(\frac{n^*-2}{N})$ and to launch the Γ_{n^*-2} which will reach full investment in less than two periods. That is, deviation yields at least $-C(\frac{n^*-1}{N}) + \sum_1^2 \delta^{t-1} B(\frac{n^*-2}{N}) + \sum_3^\infty \delta^{t-1} B(1)$, while following the proposed strategy would yield $-C(1)\delta^{(4-1)} + \sum_1^3 \delta^{t-1} B(\frac{n^*-3}{N}) + \sum_4^\infty \delta^{t-1} B(1)$. Therefore, deviating is profitable if

$$\left(\frac{n^*-2}{N}\right) - C(1)\delta^{2+d} < \sum_1^2 \delta^{t-1} B\left(\frac{n^*-2}{N}\right) + \sum_3^{2+d} \delta^{t-1} B(1) - \sum_1^{2+d} B\left(\frac{n^*-3}{N}\right) \quad (10)$$

Condition (8) rewrites conditions (9) and (10) in a general form and ensures that for any subgame with $n \geq n^* - x$ the argument just used holds. ■

Note that condition (8) never holds for $d = 0$ since the second term on the RHS vanishes and the remaining of the expression is the one analyzed at the end of Theorem 1. That is, with a delay of $n^* - n$ in any subgame we are not giving up periods with full investment benefits, $B(1)$, since by deviating the players cannot reach the full investment equilibrium earlier. On the contrary, if the delay is $n^* - n$ plus an additional delay d , following the strategy implies to give up d periods with the full investment benefit. In fact, since the second term on the RHS becomes very large for large values of d the delay before reaching an agreement is bounded. Furthermore, if the value of the full investment benefit is large enough condition (8) will always hold for any d and the maximum delay to reach a full investment outcome is $n^* - n$. More precisely, for this to hold we need $B(1)$ to check, for any $n^* \geq x > 1$:

$$B(1) > \frac{(1-\delta)\left(C\left(\frac{n^*-x+1}{N}\right) - C(1)\delta^{x+d-1}\right) + (1-\delta^{x-1+d})B\left(\frac{n^*-x}{N}\right) - (1-\delta^{x-1})B\left(\frac{n^*-x+1}{N}\right)}{(\delta^{x-1} - \delta^{x-1+d})} \quad (11)$$

To simplify the exposition, we will assume that this holds from now on.

The previous results have implicitly assumed that full investment will ultimately prevail. Nevertheless, the next two results will show that we can also have a situation where nobody ever invest (corollary 3), a result already obtained in Gale (1995) for his framework without spillovers, or even a situation where nobody continues investing even though some countries already invested (proposition 4).

Corollary 4 *For an N sufficiently large, there exists an equilibrium in which no player ever invest.*

Proof. Consider the following equilibrium strategies:

- if $n = 0$, no player ever invest.
- if $n = 1, \dots, n^* - 1$, then the remaining players invest at date $n^* - n$, if there is no deviation.
- if $n \geq n^*$, then the remaining players invest as soon as they can, that is, at date 1.

Theorem 2 has proven that the last two strategies are a SPE. It remains to be shown that the first one is also a SPE. If $n = 0$, a player can consider deviating and investing, starting then a subgame Γ_1 in which all the other players would invest in $n^* - 1$ periods (the period n^* of the original game). So, by deviating the player would face $C\left(\frac{1}{N}\right)$ in the first period and get $B\left(\frac{1}{N}\right)$ per period until n^* and $B(1)$ per period thereafter. That is, the total payoff when deviating is:

$$\left[\sum_{t=1}^{n^*-1} \delta^{t-1} B\left(\frac{1}{N}\right) - C\left(\frac{1}{N}\right) \right] + \sum_{t=n^*}^{\infty} \delta^{t-1} B(1) \quad (12)$$

On the other hand, if the player does not deviate he obtains nothing because $n = 0$ and no player ever invests and, by Assumption I, $B(0) = 0$. Therefore, for that strategy to be a

SPE (12) needs to be negative. For a large enough value of n^* we have that (12) < 0 , since by assumptions A.III the term in square bracket is negative for any value of n^* and the last term tends to zero for large values of n^* . Recalling that n^* is a non-decreasing function of N we know that proposed strategy is a SPE. That is, no player would ever invest when $n = 0$ if N is sufficiently large ($N \geq \widehat{N}$). ■

The next proposition generalizes the result in the previous corollary to a situation where some countries have already invested, for whatever reason, but the remaining countries are not interested in investing.

Proposition 5 *For N sufficiently large, perpetual non-investment is a SPE in any subgame where less than k^* players have invested, with k^* defined as the largest positive integer that checks:*

$$\left\{ \sum_1^{n^*-k} \delta^{t-1} \left[B\left(\frac{k}{N}\right) - B\left(\frac{k-1}{N}\right) \right] - C\left(\frac{k}{N}\right) \right\} + \sum_{n^*-k+1}^{\infty} \delta^{t-1} \left[B(1) - B\left(\frac{k-1}{N}\right) \right] < 0 \quad (13)$$

Proof. Consider the following strategy (and call it strategy S from now on):

- if $n < k^*$ nobody else invests.
- if $n = k^*, \dots, n^* - 1$, then the remaining players invest at date $n^* - n$, if there is no

deviation.

- if $n \geq n^*$, then the remaining players invest as soon as they can, that is, at date 1.

The two last parts of this strategy form a SPE as shown in theorem 2. Player $k^* - 1$ can consider deviating and investing, starting then a subgame Γ_{k^*} in which all the other players

would invest in $n^* - k^*$ periods (the period $n^* - k^* + 1$ of the original game). So, by deviating the player would have a cost of $C\left(\frac{k}{N}\right)$ and get $B\left(\frac{k}{N}\right)$ per period until $n^* - k$ and $B(1)$ per period thereafter. That is, the total payoff when deviating is:

$$\left[\sum_{t=1}^{n^*-k^*} \delta^{t-1} B\left(\frac{k}{N}\right) - C\left(\frac{k}{N}\right) \right] + \sum_{n^*-k^*+1}^{\infty} \delta^{t-1} B(1)$$

On the other hand, if the player does not deviate he obtains $B\left(\frac{k-1}{N}\right)$ for ever. Therefore, for that strategy to be a SPE (13) needs to be negative. For a large enough value of n^* the above expression is true at least for some values of k^* . Corrollay 3 has shown that it will always hold for $k^* = 1$, but more generally we know that the term in curly-brackets is negative for any $k^* < n^*$ because of the definition of n^* , while the second term gets smaller the larger n^* and the smaller k^* . Thus, it may also hold for $1 < k^* < n^*$. Futhermore, if this holds for a given k^* and a given N , or its corresponding n^* , it will also hold for any any $k < k^*$. ■

2.2 Coalition formation

We are now going to assume that before the implementation phase described in the previous sub-section takes place, countries can form a coalition (to simplify the exposition we will assume that only one coalition can be formed). The game has now two stages, a one-shot coalition formation stage as in Barrett (2006), followed by the implementation stage described above. We assume that if a coalition is formed with K members, it has K units of investment and that all the units have to be invested at once. If no coalition is formed the second stage of the game is exactly as the one described above. However, even if a coalition

is formed the analysis stays essentially unchanged as long as n defines the units already invested (whether by singletons or by the coalition).

We further assume that all countries act using the worst possible strategy that is a SPE, namely strategy S , or that in stage 1 of the game (coalition formation stage) all play act as if all players would play strategy S during the second stage. That is, we use a pessimistic assumption:

A.VI During the first stage of the game all players act as if all players would play strategy S during the second stage of the game (pessimistic assumption).

Our Theorem 1 has shown that if we assume short intervals we will reach full cooperation immediately. Refinements based on the assumption that players will favour Pareto efficient outcomes will also reach full cooperation immediately, since this outcome Pareto dominates all other alternatives. However, this implies solving the coordination problem almost by assumption. We are interested here in analyzing possible outcomes and possible solutions when, for whatever reasons, countries find themselves playing a SPE which will reach a Pareto sub-optimal outcome. To complicate the situation even further, we will assume that we are in a situation where $N \geq \widehat{N}$, that is, a situation where perpetual non-investment is an SPE for individual countries.

In this situation, a coalition of $K \geq n^*$ countries will always invest immediately since it can solve the coordination problem by its own (this is essentially Barrett's proposal). Nevertheless, even in the worst possible situation, where initially the number of countries is larger than \widehat{N} a coalition of less than n^* countries may be interested in investing, as the following Lemma shows:

Lemma 6 *A coalition of K countries will invest if $K \geq k^* - n$.*

Proof. A coalition of $K \geq k^* - n$ countries reverses condition (13) and is therefore interested in investing. We call the smallest coalition willing to invest \widehat{K} . ■

It remains to be seen which coalitions are stable, and the next proposition is useful for that purpose.

Proposition 7 *Assume that $N \geq \widehat{N}$, the smallest coalition willing to invest (\widehat{K}) is internally stable and externally stable if*

$$\sum_{t=1}^{n^*-\widehat{K}-1} \delta^{t-1} \left[B\left(\frac{\widehat{K}+1}{N}\right) - B\left(\frac{\widehat{K}}{N}\right) \right] + \delta^{n^*-\widehat{K}-1} \left[B(1) - B\left(\frac{\widehat{K}}{N}\right) \right] \leq C\left(\frac{\widehat{K}+1}{N}\right) - \delta^{n^*-\widehat{K}} C\left(\frac{\widehat{K}}{N}\right) \quad (14)$$

Proof. A stable coalitions needs to meet the internal stability, $\pi^S(K) \geq \pi^F(K-1)$, and the external stability, $\pi^F(K) \geq \pi^S(K+1)$, criteria. A coalition of $K = \widehat{K}$ members is always internally stable since $\pi^S(\widehat{K})$ is positive by definition while $\pi^F(\widehat{K}-1)$ is zero, since nobody ever invests (when $N \geq \widehat{N}$). External stability implies $\pi^F(\widehat{K}) - \pi^S(\widehat{K}+1) \geq 0$, with

$$\begin{aligned} \pi^F(\widehat{K}) &= \sum_{t=1}^{n^*-\widehat{K}} \delta^{t-1} B\left(\frac{\widehat{K}}{N}\right) + \sum_{t=n^*-\widehat{K}+1}^{\infty} \delta^{t-1} B(1) - \delta^{(n^*-\widehat{K}+1)-1} C\left(\frac{\widehat{K}}{N}\right) \\ \pi^S(\widehat{K}+1) &= \sum_{t=1}^{n^*-(\widehat{K}+1)} \delta^{t-1} B\left(\frac{\widehat{K}+1}{N}\right) + \sum_{t=n^*-(\widehat{K}+1)+1}^{\infty} \delta^{t-1} B(1) - C\left(\frac{\widehat{K}+1}{N}\right) \end{aligned}$$

Algebraical manipulation yields (14). ■

Although it is impossible to prove this with general functions, in general the marginal gain from having one additional member producing the benefits of cooperating immediately

(and thus provoking full cooperation one year earlier) will be lower than the increase in cost associated with investing immediately versus investing in $t = n^* - K + 1$. Thus, in general this coalition $K^* = \widehat{K}$ will also be externally stable.

To have larger coalitions stable we would need to check not only the external stability condition (which would be similar to the one just described) but also the internal stability condition, which will not anymore hold always since $\pi^F(K - 1)$ will not be zero for $K > \widehat{K}$. In fact, the same argument just used shows that the incentive to stay in the fringe may be larger than the marginal incentive of increasing a coalition that is already willing to invest. Hence, we might frequently have the situation where the smallest coalition willing to invest is the only internally and externally stable coalition. Our conjecture can be written as follows:

Conjecture 8 *The smallest coalition willing to invest is internally and externally stable. Furthermore, it is the only internally and externally stable coalition willing to invest.*

To check this conjecture, we will analyze in the next section a dynamic version of the more specific framework analyzed in Barrett (2006).

3 Application to a breakthrough technology with increasing returns

Barrett (2006) reaches highly positive results for breakthrough technologies with increasing returns to scales (what he calls X-type technologies). As a consequence, he recommends that future negotiations should focus on these type of technologies. The benefit and cost

functions that he uses are:

$$\begin{aligned}\bar{B}_i &= \bar{b}_x \left(\bar{x}_i + \sum_{j \neq i}^N \bar{x}_j \right) + \bar{b}_a \left((1 - \bar{x}_i) \bar{q}_i + \sum_{j \neq i}^N (1 - \bar{x}_j) \bar{q}_j \right), \quad \bar{x}_i, \bar{x}_j \in \{0, 1\} \\ \bar{C}_i &= \frac{\bar{c}_x}{N} \left(N - \sum_{j \neq i}^N \bar{x}_j \right) \bar{x}_i + \bar{c}_a (1 - \bar{x}_i) \frac{\bar{q}_i^2}{2}, \quad \bar{x}_i, \bar{x}_j \in \{0, 1\}\end{aligned}$$

being \bar{b}_x the marginal benefit of the technology, \bar{b}_a the marginal benefit of making an abatement, $\bar{x}_i = 0, 1$ player i 's action, with value 0 when doing an abatement and 1 when adopting the new technology and $\bar{q}_i = \frac{\bar{b}_a}{\bar{c}_a}$ the amount of abatement carried out. On the cost side, \bar{c}_x is the cost of adopting the new technology and \bar{c}_a is the marginal cost of abatement³. The player only incurs in one kind of cost depending on his action: if he abates, then $\bar{x}_i = 0$ and he only faces an abatement cost equal to $\bar{c}_a \frac{\bar{q}_i^2}{2} = \frac{\bar{b}_a^2}{2\bar{c}_a}$. If he decides to adopt the new technology rather than abating using more conventional measures, he would have a cost equal to $\bar{c}_x (1 - \bar{\alpha})$. As in our general model, the cost of adopting the new technology decreases with the number of players that have already adopted it. Multiplying and dividing by N were appropriate these functions can be written in terms of $\bar{\alpha}$:

$$\bar{B}_i = \bar{b}_x N \left(\frac{\bar{x}_i}{N} + \bar{\alpha}_{-i} \right) + \bar{b}_a \left((1 - \bar{x}_i) \bar{q}_i + \sum_{j \neq i}^N (1 - \bar{x}_j) \bar{q}_j \right), \quad \bar{x}_i, \bar{x}_j \in \{0, 1\} \quad (15)$$

$$\bar{C}_i = \bar{c}_x (1 - \bar{\alpha}) \bar{x}_i + \bar{c}_a (1 - \bar{x}_i) \frac{\bar{q}_i^2}{2}, \quad \bar{x}_i, \bar{x}_j \in \{0, 1\} \quad (16)$$

However, Barrett (2006) is a one-shot game, so that all the variables shown with a hat in (15) and (16) summarize the abatement benefits and costs from now to infinite. Since we are working in a dynamic framework, b_x, b_a, c_a, q, x and α will refer to the annualized

³See Barrett (2006) for further details on the base model

equivalent to the variables described above. Given that the benefit of abatement is linear we are implicitly neglecting the stock nature of the climate change problem and we can still show that abating each year $q_i = \frac{b_a}{c_a}$ is a dominant strategy. As before, we assume that the cost of adopting the technology is faced in the period when it is adopted and this cost is given by $C = c_x(1 - \alpha)$. Thus, in our dynamic framework the discounted payoff for player i of adopting the new technology at time s is: .

$$\begin{aligned}\pi_{i,1} &= B_{i,1} - C_{i,1} \\ B_{i,1} &= \sum_{t=1}^{\infty} \delta^{t-1} \left(b_x N \left(\frac{x_{i,t}}{N} + \alpha_{-i,t} \right) + b_a \left((1 - x_{i,t}) q_i + \sum_{j \neq i}^N (1 - x_{j,t}) q_j \right) \right) \\ C_{i,1} &= \delta^{s-1} c_x (1 - \alpha_t) x_{i,s} + \sum_{t=1}^{\infty} \delta^{t-1} \left[c_a (1 - x_{i,t}) \frac{q_i^2}{2} \right]\end{aligned}$$

$$\text{with } x_{i,t} = \begin{cases} 0 & \text{if } t < s \\ 1 & \text{if } t \geq s \end{cases} \text{ and } x_{i,t} \geq x_{i,t-1} \forall t = 1, \dots, \infty$$

The following lemma gives the conditions under which our assumptions hold with our dynamic version of Barrett's functions.

Lemma 9 *Assumptions II to V hold for the benefit and cost functions defined in (15) and*

(16) if

$$c_x \left(1 - \frac{1}{N} \right) (1 - \delta) + \frac{b_a^2}{2c_a} > b_x > \frac{b_a^2}{c_a} \quad (17)$$

Proof. Appendix A. ■

Note, however, that Assumption I does not hold for the benefit side, since in Barrett's model the country abates if it decides not to invest in the technology. Nevertheless, this does

not change our main results so far, since it essentially implies rescaling the benefit function.

Hence, if the parameter values check these inequalities, we know that for some $0 < n^* < N$:

$$\sum_{t=s}^{\infty} \delta^{s-1} \{[n^* b_x + (N - n^*) b_a q_i] - [(n^* - 1) b_x + (N - (n^* - 1)) b_a q_i]\} = c_x \left(1 - \frac{n^*}{N}\right) - \sum_{t=s}^{\infty} c_a \frac{q_i^2}{2}$$

or

$$n^* = \left(1 - \frac{b_x - \frac{b_a^2}{2c_a}}{c_x(1 - \delta)}\right) N$$

With a set of parameters satisfying (17), Theorem 2, Corollary 4 and Proposition 5 hold.

Thus, for N large enough there exists a SPE in which no player ever invest. In particular, with the functional forms used in this section the latter strategy is a SPE if the benefit of not investing

$$\sum_{t=1}^{\infty} \delta^{t-1} \left(b_a N q_i - c_a \frac{q_i^2}{2}\right) = \sum_{t=1}^{\infty} \delta^{t-1} \frac{b_a^2}{c_a} \left(N - \frac{1}{2}\right) \quad (18)$$

is larger than the benefit of investing

$$\sum_{t=1}^{n^*-1} \delta^{t-1} \left(b_x + b_a (N - 1) \frac{b_a}{c_a}\right) - c_x \left(1 - \frac{1}{N}\right) + \sum_{t=n^*}^{\infty} \delta^{t-1} b_x N \quad (19)$$

That is, we need (19) < (18):

$$\delta^{n^*-1} (N - 1) \left(b_x - \frac{b_a^2}{c_a}\right) + \left(b_x + \frac{1}{N} (\delta - 1) (N - 1) c_x - \frac{b_a^2}{2c_a}\right) < 0 \quad (20)$$

Equation (20) cannot be solved analytically for N , and we therefore resort to simulations with values of the parameters satisfying (17). Figure 1 plots the value of the LHS of (20) for the parameters shown. With these parameters, for $N > 51 = \hat{N}$ the value of the LHS of

(20) becomes negative. Hence, (20) holds and not investing when nobody else has invested ($n = 0$) is a SPE.

[Figure 1]

3.1 The role of treaty design

The last section has shown that, for the parameters used in Figure 2, if there are more than $\hat{N} = 51$ countries no country will ever invest. Can a treaty improve this situation? As stated above, usually treaties are modeled in this literature (Barrett, 2006) assuming that all the countries form a coalition that from now on maximizes its joint welfare. Suppose $N = 100$ and that countries can get together and form a single coalition with K members. Then, for each one of the members of the coalition not investing implies to follow the optimal abatement policy, which yields the following net benefits (if $K < n^*$):

$$S^N(K) = \sum_{t=1}^{\infty} \delta^{t-1} \left(K b_a q_K + (N - K) b_a q_i - c_a \frac{q_K^2}{2} \right) = \frac{b_a^2}{(1 - \delta) c_a} \left(N + \frac{K^2}{2} - K \right) \quad (21)$$

with $q_i = \frac{b_a}{c_a}$ and $q_K = \frac{K b_a}{c_a}$. That is, as in the canonical model the coalition would abate more than any singleton in the eventually of choosing to abate instead of investing in the new technology.

The net benefit of investing in the new technology the K units that the coalition has

available is (if $K < n^*$)

$$\begin{aligned}
S^I(K) &= \sum_{t=1}^{n^*-K} \delta^{t-1} \left(K b_x + (N-K) \frac{b_a^2}{c_a} \right) - c_x \left(1 - \frac{K}{N} \right) + \sum_{t=n^*-K+1}^{\infty} \delta^{t-1} b_x N \quad (22) \\
&= \frac{1 - \delta^{n^*-K}}{1 - \delta} \left(K b_x + (N-K) \frac{b_a^2}{c_a} \right) + \frac{\delta^{n^*-K}}{1 - \delta} b_x N - c_x \left(1 - \frac{K}{N} \right)
\end{aligned}$$

since the remaining players will invest after $n^* - K$ periods, or in period $n^* - K + 1$ of the original game. That is, investing is optimal when $S^N(K) < S^I(K)$, or:

$$\delta^{n^*-K} (N-K) \left(b_x - \frac{b_a^2}{c_a} \right) + \left(K b_x + \frac{1}{N} (\delta - 1) (N-K) c_x - \frac{K^2 b_a^2}{2c_a} \right) > 0 \quad (23)$$

Computing this equation, for $N = 100$, figure 2 shows that only coalitions between 4 and 9 members would invest.

[Figure 2]

But, which coalitions would be stable? As indicated in the previous section, a stable coalitions needs to meet the internal stability, $\pi^S(K^*) \geq \pi^F(K^* - 1)$, and the external stability, $\pi^F(K^*) \geq \pi^S(K^* + 1)$, criteria. To compute this we need to take into account that:

$$\pi^S(K) = \begin{cases} S^I(K) & \text{if } S^N(K) < S^I(K) \\ S^N(K) & \text{if } S^N(K) \geq S^I(K) \end{cases}$$

Assuming that we are in the part where singletons are not interested in investing (i.e.

that $N > \widehat{N}$) the net benefit of those not investing is

$$\begin{aligned} F^I(K) &= \sum_{t=1}^{n^*-K} \delta^{t-1} \left(K b_x + (N-K) \frac{b_a^2}{c_a} \right) + \sum_{t=n^*-K+1}^{\infty} \delta^{t-1} b_x N - \sum_{t=1}^{n^*-K} \delta^{t-1} c_a \frac{q_i^2}{2} + \delta^{(n^*-K+1)-1} c_x \\ &= \frac{1 - \delta^{n^*-K}}{1 - \delta} \left(K b_x + (N-K) \frac{b_a^2}{c_a} \right) + \frac{\delta^{n^*-K}}{1 - \delta} b_x N - \frac{1 - \delta^{n^*-K}}{1 - \delta} \frac{b_a^2}{2c_a} - \delta^{n^*-K} c_x \end{aligned}$$

if the members of the coalition invest and

$$\begin{aligned} F^N(K) &= \sum_{t=1}^{\infty} \delta^{t-1} \left(K b_a q_K + (N-K) b_a q_i - c_a \frac{q_i^2}{2} \right) \\ &= \frac{b_a^2}{(1 - \delta) c_a} \left(N + K^2 - K - \frac{1}{2} \right) \end{aligned}$$

if the members of the coalition do not invest. That is, we have that

$$\pi^F(K) = \begin{cases} F^I(K) & \text{if } S^N(K) < S^I(K) \\ F^N(K) & \text{if } S^N(K) \geq S^I(K) \end{cases}$$

Computing these values shows that, as in the canonical model, a coalitions of 2 and 3 members is internally and externally stable but only abates. Thus, with these coalitions the new technology may never be implemented. However, we also find that a coalition with 4 members is internally and externally stable and that this is the only internally and externally stable coalition willing to invest (confirming Conjecture 8 for this particular case). In addition, this coalition invests in the new technology and therefore launches the process that will eventually allow universal adoption of the new technology.

4 Conclusions

This paper has analyzed a monotone game over the implementation of a pure public good. After discussing the general model, we have applied this model to the analysis of a breakthrough technology with increasing returns. Our results suggest that if the number of countries (agents) is small, the technology will be adopted without the need of any treaty. Nevertheless, the technology will not necessarily be adopted immediately and the potential delay will be proportional to the number of countries. On the contrary, if the number of countries is large the technology may never be adopted.

Accepting that a treaty can be modeled by assuming that all the countries signing an agreement maximize their joint welfare, we have shown that even a small coalition of countries would be interested in adopting the technology and launching the process that will ultimately lead to the universal adoption of the technology.

References

- [1] Barrett, S. (1994), Self-Enforcing International Environmental Agreements, *Oxford Economic Papers*, New series, 46, 878-894.
- [2] Barrett, S. (2006), Climate treaties and "breakthrough" technologies, *American Economic Review*, 96, 2, 22-25.
- [3] Carraro, C. and Siniscalco, D. (1993), Strategies for the International Protection of the Environment, *Journal of Public Economics*, 52, 309-328.
- [4] Gale, D. (1995), Dynamic coordination games, *Economic Theory*, 5, 1-18.

- [5] Gale, D. (2001), Monotone Games with Positive Spillovers, *Games and Economic Behavior*, 37, 295-320
- [6] Hoel, M. and de Zeeuw, A. (2009), *Can a focus on breakthrough technologies improve performance of international environmental agreements?*, National Bureau of Economic Research, Working Paper 15043.

A Assumptions I to V with our dynamic version of Barrett's functions

With our dynamic version of Barrett's (2006) X-type functions Assumption II simplifies to

$$\sum_{s=1}^{\infty} \delta^{s-1} (b_x - b_a q_i) > - \sum_{s=1}^{\infty} c_a \frac{q_i^2}{2}$$

or:

$$b_x > \frac{b_a^2}{2c_a} \tag{24}$$

Assumption III now simplifies to

$$b_x < c_x \left(1 - \frac{1}{N}\right) (1 - \delta) + \frac{b_a^2}{2c_a}$$

Then, combining (24) with this inequality we need:

$$c_x \left(1 - \frac{1}{N}\right) (1 - \delta) + \frac{b_a^2}{2c_a} > b_x > \frac{b_a^2}{2c_a}$$

These restrictions on the parameters are equivalent to those assumed by Barrett (2006), differing only in the dynamic character of our model.

Assumption IV also holds since $B_i = b_x N \left(\frac{x_i}{N} + \alpha_{-i}\right) + b_a \left((1 - x_i) q_i + \sum_{j \neq i}^N (1 - x_j) q_j\right)$, $x_i, x_j \in \{0, 1\}$ is continuous and non-negative for reasonable values of the parameters. The incremental benefit for any player investing when n other players have already invested, for

any $0 \leq n < N$, is $B(n+1) - B(n) = b_x - b_a q_i$. To have $B(\alpha_t)$ increasing in α_t /number of player investing, we need:

$$b_x > \frac{b_a^2}{c_a}$$

Assumption V also holds. In our dynamic version of Barrett's model $C = c_x(1 - \alpha)$ for any t , so it is continuous and clearly decreasing on α . Finally, Assumption I holds for C and just needs a rescaling for B .

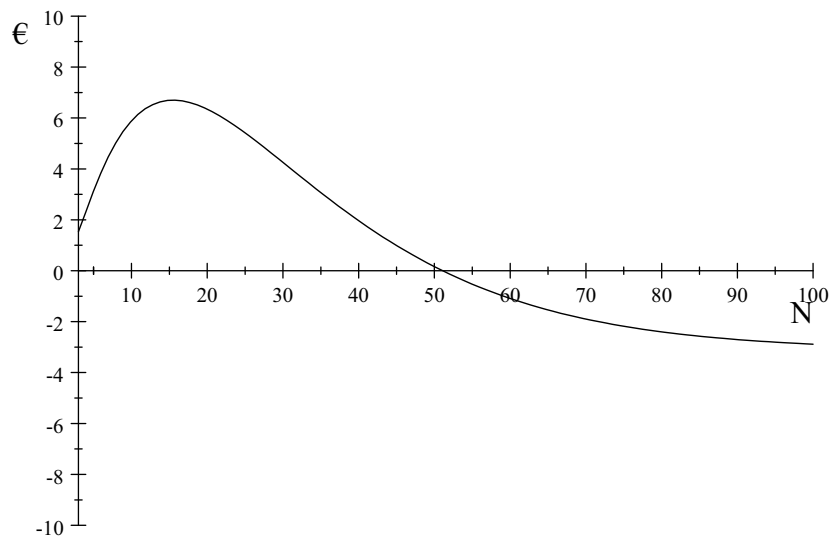


Figure 1. Marginal payoff of investing for the first mover (parameters: $b_a = 1$, $b_x = 2$, $c_a = 3$, $c_x = 50$, $\delta = 0.9$)

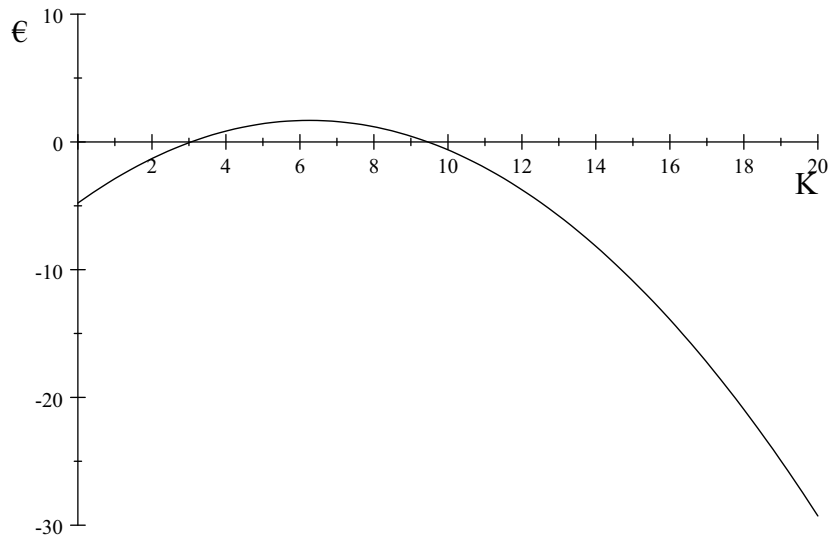


Figure 2. Marginal payoff of investing as the first mover for a coalition of K members with $N=100$ (parameters: $b_a = 1$, $b_x = 2$, $c_a = 3$, $c_x = 50$, $\delta = 0.9$)