Why and when coalitions split? An alternative analytical approach with an application to environmental agreements

Raouf Boucekkine, Carmen Camacho, Weihua Ruan, Benteng Zou

PARIS SCHOOL OF ECONOMICS
ECOLE D'ECONOMIE DE PARIS

WORKING PAPER N° 2022 – 16

Why and when coalitions split? An alternative analytical approach with an application to environmental agreements

Raouf Boucekkine
Carmen Camacho
Weihua Ruan
Benteng Zou

JEL Codes: C61, C73, D71.
Keywords: Coalition splitting; Environmental agreements; Constitutional vs technological heterogeneity; Differential games; Multistage optimal control.

agence nationale de la recherche
AU SERVICE DE LA SCIENCE

Why and when coalitions split? An alternative analytical approach with an application to environmental agreements*

Raouf Boucekkine† Carmen Camacho‡ Weihua Ruan§
Benteng Zou¶

May 20, 2022

Abstract

We use a parsimonious two-stage differential game setting where the duration of the first stage, the coalition stage, depends on the will of a particular player to leave the coalition through an explicit timing variable. By specializing in a standard linear-quadratic environmental model augmented with a minimal constitutional setting for the coalition (payoff share parameter), we are able to analytically extract several nontrivial findings. Three key aspects drive the results: the technological gap as an indicator of heterogeneity across players, the constitution of the coalition and the intensity of the public bad (here, the pollution damage). We provide with a full analytical solution to the two-stage differential game. In particular, we characterize the intermediate parametric cases leading to optimal finite time splitting. A key characteristic of these finite-time-lived coalitions is the requirement of the payoff share accruing to the splitting country to be large enough. Incidentally, our two-stage differential game setting reaches the conclusion that splitting countries are precisely those which use to benefit the most from the coalition. Constraining the payoff share to be low by Constitution may lead to optimal everlasting coalitions only provided initial pollution is high enough, which may cover the emergency cases we are witnessing nowadays.

Keywords: Coalition splitting; environmental agreements; constitutional vs technological heterogeneity; differential games; multistage optimal control.
JEL classification: C61, C73, D71.

*In memory of Thierry Bréchet.

†Rennes School of Business, France. E-mail: raouf.boucekkine@rennes-sb.com
‡Paris School of Economics and CNRS, France. E-mail: carmen.camacho@psemail.eu.
§Purdue University Northwest, USA. E-mail: Wruan@pnw.edu
¶DEM, University of Luxembourg, Luxembourg. E-mail: benteng.zou@uni.lu.

1 Introduction

Recently, numerous withdrawals of countries from international organizations and agreements have been witnessed: (1) the most recent happens on July 7, 2020, the Trump administration formally notified the United Nations that it is pulling out of the World Health Organization, which effective as of July 6th, 2021;[1] (2) during the same Trump presidency period, on June 1, 2017, President Trump announced that the U.S. would cease all participation in the 2015 Paris Agreement on climate change mitigation until some fair conditions to the U.S.A could be negotiated; (3) the United Kingdom withdrew from the European Union on January 31, 2020; (4) Canada withdrew from Kyoto Protocol on December 13, 2011... An analogous trend has been taking place within and between allied companies with a noticeable impact social and economic/financial effects. For example, on October 2014, technology giant Hewlett-Packard, known as HP, splits itself into two separate companies: separating its computer and printer businesses from its faster-growing corporate hardware and services operations, causing the dismissal of about 5,000 jobs as part of its turnaround plan. On July 2015, PayPal spins off from eBay and this split benefits both eBay's marketplace business by letting it accept different forms of electronic payment and also gives PayPal more autonomy to work with other potential partners, such as Amazon or Alibaba.

Obviously, this kind of phenomena has drawn much attention in the economic literature. A bunch of papers investigating the impact of Brexit (Sampson, 2017; Latorre et al., 2020; the special issue of the Oxford Review of Economic Policy, vol 33, 2017; etc.), or the economic consequences of U.S. withdrawal from the Kyoto Protocol and Paris agreement (Bucher et al., 2002; Dai et al., 2017; Nong and Siriwardana, 2018; ...) have been already published.[2] The broader questions of the design of international agreements (binding or not) and the emergence and stability of coalitions have become a very active research lines since the early 90s in environmental economics. The game-theoretical settings proposed have been quite diverse ranging from cooperative to non-cooperative, from static to dynamic through repeated games, and often including some interesting procedural ingredients, typically on enforceability of the agreements. See surveys in Bréchet et al. (2011)

[1] But at the same time, Joe Biden, who challenged Donald Trump in the November 2020 presidential election, tweeted: "On my first day as President, I will rejoin the WHO and restore our leadership on the world stage, " which he eventually did, although not on his first presidential day

[2] See also the empirical study of Mayer et al (2019) on the cost of being non-EU, and the general theoretical investigation of Gancia et al (2020) on the gain of being in some economic unions and partnerships.

and Calvo and Rubio (2013). We shall focus on this stream of the economic literature as our working model is based on benchmark models borrowed from this stream.

Regarding this abundant literature, the coalition splitting problem has been addressed within two different conceptual settings. The first and probably more natural one is based on the theory of coalition stability as explained in Tulkens (1998) and surveyed by Bréchet et al. (2011). The theory is essentially anchored in the cooperative games literature[3]. It relies on two concepts of coalition stability, the very known core-stability and the less known internal-external stability concepts. The main difference between the two concepts is basically the size of coalitions playing the game: the core-stability concept focuses on strategies chosen by the members of the grand coalition (including all countries) while internal-external stability is concerned by the set of strategies chosen by any coalitions of any size, and practically considers the benefits for each country of being inside or outside these coalitions. Beside the many interesting purely theoretical questions that this dichotomy may suggest, the value-added of the two concepts in terms of potential policy recommendations are generally different as the second concept is inherently more likely to point at the most viable (stable) coalition size. This is nicely illustrated in Bréchet et al. (2011) on the climate agreements.

Most of the literature on environmental agreements is however anchored in the traditional Nash non-cooperative theory with individual strategies (with a Pareto-like criterion to evaluate efficiency). The basic ingredients of this stream are well identified in Calvo and Rubio's survey (2013). Typically in this literature, n given countries (possibly non-identical) have to decide "individually" whether to join an environmental agreement to jointly fight the public bad. Of course, having a binding vs a non-binding agreement is important. But not that much: it's indeed unclear whether a binding agreement will prevent players from free-riding and engaging in the intended coalitions. Insightful work has been done on these issues and many other related ones in a variety of frameworks ranging from multistage games *à la* Carraro and Siniscalco (1993) to dynamic games (Hoel, 1993; Xepapadeas, 1995; Dutta and Radner, 2009) through the incomplete contracting-based games *à la* Harstad (2012) and Battaglini and Harstad (2016).

In this paper, we take a different approach originating from a different research question. Consider a coalition of countries initially maximizing a given joint payoff subject to a pub-

[3]Recall that non-cooperative games consider strategies of individual players consistently with the Nash equilibrium concept, while cooperative games extend the analysis to the strategies chosen jointly by groups of players.

lic bad (or good). The coalition is based on constitutional rules, some related to sharing quotas (of benefits and costs of the coalition) and other possibly on enforceability. Also, the coalition may entail large heterogeneities across players: technological, demographic or geographic notably. Given all these ingredients, under which conditions a country initially belonging to this coalition may eventually split at a finite date, and **when**? What are the determinants (Constitution, technology,...etc) of splitting and of the duration of the coalition? The literature surveyed above is not focused on this question, and the timing issues are very seldom explicitly treated. In general, it's implicitly determined at the games equilibria (see Battaglini and Harstadt, 2016). In our framework, the splitting time is an explicit optimal control in the hands of any member of initial coalitions.

Beside the timing question, our seemingly broad research question is accurate, the main objective pursued being the identification of the constitutional vs countries' heterogeneities factors which *per se* lead to coalition splitting, in the absence of any aggregate or country-specific exogenous shock. To tackle it, we shall consider the following theoretical setting. Initially players (countries) agree to manage cooperatively the public bad or good, the stock of pollution in our working example. As a shortcut to the constitutional aspects of the coalition, we assume that each country enters the coalition with a given fixed share of the (intertemporal) payoff of the coalition. We could have added an additional ingredient formalizing punishment in case of 'cheating' or, simply, withdrawal from the initial coalition. We don't ease the algebra, the methodology is unaltered. We also assume that this constitutional weight is independent of the technological level of the countries, in order to disentangle the implications of pure constitutional vs technological/economic effects. More importantly, we allow indeed for constitutional and technological heterogeneities across countries, which is the most interesting ingredient from the economic viewpoint.

We further study under which conditions a given country of the initial coalition may eventually split at a given finite time T. If a country splits, a non-cooperative game sets in between the country and the group of countries remaining in the coalition. Within a full linear-quadratic model, we characterize the optimal affine Markovian subgame perfect strategies for a given split time T. We later solve for the whole sequence (starting with the initial cooperative game phase) and uncover the conditions under which splitting occurs at finite time. In particular, we study in depth the determinants of the duration of the coalition with particular attention paid to the role of technological vs constitutional heterogeneity across players.

With respect to the timing questions, and as stated in Boucekkine et al (2020), these issues are seldom explicitly addressed though the vast majority of the models developed in the related literature are dynamic. An exception is Battaglini and Harstad (2016) but in a completely different framework. In particular, the latter builds on renegociation games within an incomplete contracting game, not differential games.[4] We shall build a parsimonious multistage differential game framework to deal with our research questions analytically. The multistage nature of the set-up is natural as the game starts with the coalitional structure before a possible move to a second stage if the splitting decision is taken at finite time by an initial member. But in contrast to Carraro and Siniscalco (1993), the move to the second stage involves an explicit timing decision, that ultimately (endogenously) changes the nature of the game to a Nash differential game if the second stage is accessed. Last but not least, we don't allow for renegotiation contrary to the games *à la* Harstad: usually environmental agreements do not allow for renegotiations over a given period of time (see for example, Hattori, 2001); here, we simply assume that this period is infinite.[5]

Technically, our analytical approach requires the tools of multistage optimal control together with the typical techniques needed to solve differential games There exist an increasing number of papers using multistage optimal control to characterize optimal/ equilibrium regime transitions and the inherent timings (Boucekkine et al., 2013; Moser et al., 2014; Saglam, 2011; Zampolli et al. 2016; etc.). Commonly, all these studies rely on the seminal exploration of Tomiyama (1984). In contrast, much fewer papers merging multi-stage optimal control and dynamic games have come out (see for example Boucekkine et al., 2011, or Boucekkine et al., 2022). We shall show how the latter avenue can also be taken safely in our paper, though analytical results may only be secured in the linear-quadratic case.

Three key aspects drive the results: the technological gap as an indicator of heterogeneity across players, the Constitution of the coalition (captured by a single parameter, the payoff share accruing to countries under coalitions) and the pollution damage. Thanks

[4] Incidentally, Harstad (2012) has heavily criticized the use of differential games as often implying bang-bang solutions due the typical linear-quadratic assumptions adopted for tractability. We shall see that this argument is not valid for the class of differential games we introduce despite linear-quadratic assumptions, because of the explicit splitting time control variable.

[5] Introducing an exogenous no-renegotiation period into the model is not in itself a complicated matter but it will make the analysis much more cumbersome as an addition and potentially very tedious discussion would be needed to position the optimal splitting times with respect to the no-renegotiation time. We omit this complication here.

to this parsimonious specifications, we are able to provide with a fully analytical solution to the two-stage differential game under scrutiny. We do cover all the set of parameterizations taken by the three indicators listed above, which results in a highly nontrivial mathematical analysis (despite parsimony). In particular, we characterize the intermediate parametric cases leading to optimal finite time splitting. A key characteristic of these finite-time-lived coalitions is the requirement of the payoff share accruing to the splitting country to be large enough. Incidentally, our two-stage differential game setting reaches the conclusion that splitting countries are precisely those which use to benefit the most from the coalition. Constraining the payoff share to be low by Constitution may lead to optimal everlasting coalitions only provided initial pollution is high enough, which may cover the emergency cases we are witnessing nowadays.

The paper is organized as follows. Section 2 briefly presents the general specification of our environmental game-theoretical setting. Section 3 fully analyzes a specialized linear-quadratic version of the game, providing in particular the optimal players' strategies for given splitting time. Section 4 characterizes the existence of an optimal (interior) splitting time, discusses its constitutional and technological drivers and delivers some policy insights. Section 5 concludes.

2 The model of optimal splitting

Suppose there is one bloc of players, such as a confederation of regions or a pro-environmental coalition like those which resulted from the Kyoto Protocol or the Paris climate Agreement. Suppose that one of the bloc's players, named player i, would eventually like to quit the bloc at some future date T. The rest of the bloc is named as player J. The two players, i and J, share a common variable, $y \in [0, Y]$.

When players act as one bloc, players i and J choose jointly the level of variables $x_i, x_J \in [0, X] \subset [0, +\infty)$, which provide them with a joint utility or payoff. Their choices for x_i and x_J may increase or lower the level of y, which may induce a loss or gain in utility. In this sense, y may be a public good or a public bad. To fix ideas, we will consider that the players' actions increase the level of y, resulting in a drop in utility. This mimics the typical public bad (pollution) model in environmental economics, which will be our workhorse along this paper. In our model, we assume that at time 0, players play cooperatively until time T, when player i decides to quit the bloc. Note that at time T, player J may also switch her strategy in response. We shall concentrate on the Markovian perfect equilibria of the game.

Coming back to our workhorse pollution model, suppose there exists a unique final good, which requires only a polluting resource as input. With a quantity of pollution x_j, $j \in \{i, J\}$, both players j produce the final good, y. The consumption of this final good provides players with utility, but at the same time the production process increases the level of CO_2 emissions, which we assume as usual equal to production, y. Obviously, the level of CO_2 affects both players. In the end, player j can obtain utility directly from the consumption of x_j, but she also suffers from pollution, since y brings disutility.

Initially, the objective of the players in the bloc is to maximize joint overall welfare or payoff, defined as

$$\max_{x_i, x_J} W(\infty) = \int_0^{+\infty} e^{-rt} \left[u_i(x_i) + u_J(x_J) - c_i(y) - c_J(y) \right] dt, \tag{1}$$

where r is the time discount rate, u_i and u_J are, respectively, utility functions of player i and J, which are strictly increasing and concave; and $c_i(y), c_J(y)$ are their respective disutility due to pollution, which are strictly increase and convex. The objective function is simply the aggregate payoff of the two players. We shall discuss just below how the optimal payoff is shared across players when we come to the constitutional bases of the initial coalition.

Finally, decisions are subject to the dynamic constraint:

$$\dot{y}(t) = f(x_i, x_J, y) = x_i + x_J - \delta y(t), \quad y(0) = y_0 \text{ given}, \tag{2}$$

and $\delta \in [0, 1]$ is the depreciation rate. In our example, y stands for CO_2 emissions, so that δ would stand for the natural reabsorption rate of CO_2 in the atmosphere.

We get now to the **constitutional aspects** of the coalition. Essential aspects are sharing rules (of the aggregate payoff) and penalties in case of splitting. As argued in the Introduction, we shall focus on the first aspect. The second one is abundantly studied in the literature (see again Calvo and Rubio, 2013), and it could be trivially introduced as an additional cost for player i to pay at the splitting time T (with the corresponding algebraic complications). We simplify here the exposition and focus on the sharing rule. Concretely, we suppose that player i's share in the total payoff is $\alpha \in (0, 1)$ and the remaining share, $1-\alpha$, belongs to the rest of the bloc, i.e., welfare of players i and J are

$$W_i = \alpha W(\infty) \text{ and } W_J = (1-\alpha)W(\infty).$$

Two points are worth doing at this stage, both deriving from the fact that we consider a fixed share α. First of all, while α is defined as a fraction of the intertemporal and discounted payoff, it indeed applies at any period of time since it's constant. We can therefore interpret it also as an instantaneous share. This said, the payoff to be shared in our game-theoretic framework is the intertemporal one: when splitting is an option, player i will consider the share of the intertemporal payoff from the start of the coalition to its (potential) end at date T, that is $\alpha W(T)$ where $W(T) = \int_0^T e^{-rt}\left[u_i(x_i) + u_J(x_J) - c_i(y) - c_J(y)\right]dt$. Second, beside this technical point, one would inquire about the particular meaning of such a sharing in a public bad problem like this one. One would be naturally tempted to bring this aspect closer to the standard literature of environmental agreements where the enforcement of a Pareto efficiency criterion would require transfers from certain countries to others (see the early contribution of Tahvonen, 1994). However, this is precisely what we don't do in this model, the weight α is by no way a Pareto weight: it is *stricto sensu* a constitutional parameter, it's fixed initially with the birth of the coalition according to the initial political, demographic or economic relative powers of the members.

Of course, α would be generally dependent on the characteristics of each country member of the coalition, that is on the shapes of the national preferences and technologies. But in reality α also depends on more complex characteristics like global and regional history and geography and the resulting regional and global geopolitics which can hardly be recovered unequivocally from technological differentials or cultural differences. In this paper, we define the constitutional rule as being independent from the $u_j(.)$ and $c_j(.)$ to clearly discriminate between the constitutional foundations of the coalition and the more purely technological diversity. We consider this case as the natural benchmark to explore in depth.

Moreover, we assume that renegotiation (of α) is impossible or too costly, which is more than realistic.[6]. Note that this could be one of the reasons why player i could decide to quit the bloc at some future date T and play non-cooperatively. We shall see that this outcome is far from automatic and it is only obtained under highly nontirival conditions.

[6] One way to avoid player i quitting the bloc is allowing renegotiation such that α is a function of contribution, $\alpha = \alpha(x_i)$, or ratio of contribution: $\alpha = \alpha(x_i, x_J)$. We shall show here that the analysis of our benchmark with α constant is far nontrivial and deserves a separate and deep analysis.

We now move to some preliminary technical considerations. If player i quits the bloc at time T, then as already said, she obtains a share α of overall welfare until time T. From time T onwards, player i's objective becomes

$$W_{i,II} = \max_{x_i} \int_T^{+\infty} e^{-rt} \left[u_i(x_i) - c_i(y)\right] dt, \tag{3}$$

and player J faces

$$W_{J,II} = \max_{x_J} \int_T^{+\infty} e^{-rt} \left[u_J(x_J) - c_J(y)\right] dt, \tag{4}$$

subject to the same state equation:

$$\dot{y}(t) = f(x_i, x_J, y) = x_i + x_J - \delta y(t), \quad t \geq T, \tag{5}$$

where initial condition $y(T)$ comes from the outcome of the first period.

The optimal switching time for player i is defined as

$$\max_T \left(\alpha W(T) + \int_T^{+\infty} e^{-rt} \left[u_i(x_i) - c_i(y)\right] dt \right) = \max_T \left(\alpha W(T) + W_{i,II} \right), \tag{6}$$

where $W(T)$ is the same integral in (1) but over time interval $[0, T]$:

$$\int_0^T e^{-rt} \left[u_i(x_i) + u_J(x_J) - c_i(y) - c_J(y)\right] dt.$$

Intuitively, the first term in (6) is non-decreasing in T. That is, the longer the time period, the higher is the social welfare, otherwise, if the first term is decreasing in T, then joining the coalition is never welfare enhancing for player i, and T should be 0. Similarly, the second term in (6) is non-increasing with T. Otherwise, if the second term was also increasing in T, then it would always be optimal to set $T = +\infty$. Obviously, the precise optimal choice of T relies on the strategy space after the splitting. Depending on the parameter set, the optimal choice of T can be 0, ∞ or take any other finite value between 0 and ∞. If the optimal choice is $T = +\infty$ then quitting a coalition will never be optimal.

The interior optimal switching time T is given by the solution of

$$\alpha \frac{dW(T)}{dT} + \frac{dW_{i,II}}{dT} = 0,$$

provided the second order optimality condition

$$\alpha \frac{d^2W(T)}{dT^2} + \frac{d^2W_{i,II}}{dT^2} < 0$$

holds.

In the next section with linear-quadratic model, we pay special attention to the situation where $T = +\infty$, that is the situation in which the coalition will continue forever. Here, let us assume first that there exists a unique interior solution, that is , that $0 < T < \infty$. Then, invoking the implicit function theorem we obtain that

$$\frac{\partial T}{\partial \alpha} = -\frac{\frac{dW(T)}{dT}}{\alpha \frac{d^2W(T)}{dT^2} + \frac{d^2W_{i,II}}{dT^2}} > 0.$$

In other words, if player i is the dominating player in the bloc. i.e., occupying a relatively larger share from the aggregate welfare, then she will quit the bloc later. By the same token, a small player would rather quit the bloc earlier to be free from commitments. . It is worth mentioning that time T should not be considered as a truncated terminal time, rather it is the time at which the splitting happens. Different from most of the optimal switching literature, under the current setting, before and after time T, players change: before time T, the bloc makes choice for the bloc as a single player while after T there are two competing players. Thus special care should be taken when employing the usual necessary optimal switching conditions at T. This difficulties come mainly from the choice of different strategic spaces after the splitting. In the following sections, we introduce some of these cases.

3 The linear-quadratic case two-stage differential game

To hopefully solve in a fully analytical way the above problem, we resort as it is traditional in differential games to linear-quadratic functional forms (see Dockner and Van Long,1993; Dockner et al, 2000; Bertinelli et al, 2014; etc). In this section, we focus on the optimal strategies at each stage of the game for given splitting time, Section 4 will adress the optimal splitting time issues.

In our linear-quadratic setting, the utility functions are given by

$$u_i(x_i) = a_i x_i - \frac{x_i^2}{2}, \quad u_J(x_J) = a_J x_J - \frac{x_J^2}{2},$$

and the pollution damage functions are

$$c_j(y) = \frac{by^2}{2}, \quad j = i, J.$$

In other words, regardless the development level, the pollution damage is the same for both players for simplicity. x_j can be considered as the pollution emission of player j in order to produce final consumption goods, and a_j as the efficiency parameter which converts pollution into the consumption good. Thus, a higher a_j indicates a more advanced economy, meaning that it can generate more of the consumption good from the same unit of pollution.

3.1 Optimal strategies in the cooperative stage (before T)

The joint payoff function is

$$\max_{x_i, x_J} W(\infty) = \int_0^{+\infty} e^{-rt} \left[a_i x_i + a_J x_J - \frac{x_i^2 + x_J^2}{2} - by^2 \right] dt, \tag{7}$$

subject to the following state equation:

$$\dot{y}(t) = x_i + x_J - \delta y(t), \quad y(0) = y_0 \text{ given}, \tag{8}$$

and as mentioned above $\delta \in [0, 1]$ is the depreciation rate. In our example, y stands for CO_2 emissions, so that δ would stand for the natural reabsorption rate of CO_2 in the atmosphere.

The first best solution for $t \in [0, +\infty)$ is obtained from the bloc's optimal control problem with objective function (7) and subject to the state equation (8). We can readily summarize the main results of this part of the problem in the following proposition.

Proposition 1 *For any positive constants b, r, δ, then for any state trajectory $y(t)$, the* *optimal choices for player i and J are*

$$x_j^*(y) = a_j + B + Cy, \quad j = i, J,$$

where

$$C = \frac{r + 2\delta - \sqrt{(r+2\delta)^2 + 16b}}{4}(< 0), \quad B = \frac{(a_i + a_J)C}{r + \delta - 2C}(< 0).$$

The optimal trajectory of state is: $\forall t \geq 0$,

$$y(t) = (y_0 - y^*)e^{(2C-\delta)t} + y^*$$

where y^* is the optimal asymptotically stable long-run steady state and given by

$$y^* = \frac{a_i + a_J + 2B}{\delta - 2C}(> 0).$$

The long-run steady state y^* depends on all the parameters, especially the sum of the technology levels, a_i and a_J. A higher technology level, which translates into higher consumption, yields higher long-run pollution accumulation. Consistently with the standard linear-quadratic model considered, the convergence speed, $(2C - \delta)$, is independent of the technology levels, it rather depends on time preference r, the unit damage of pollution, b, and nature's regeneration rate δ.

We notice that the two players' aggregate consumption at the long-run steady state y^* is always positive:

$$x_i^*(y^*) + x_J^*(y^*) = a_i + a_J + 2B + 2Cy^* = \frac{\delta}{\delta - 2C}(a_i + a_J + 2B) > 0,$$

where the last inequality comes from the fact that $a_i + a_J + 2B > 0$. Furthermore, actually the aggregate consumption is always positive along the optimal trajectory path, that is, $\forall y$ and $\forall r, \delta, a_i, a_J, b > 0$, it follows

$$\begin{aligned} x_i^*(y) + x_J^*(y) &= a_i + a_J + 2B + 2Cy = a_i + a_J + 2B + 2C[(y_0 - y^*)e^{(2C-\delta)t} + y^*] \\ &= x_i^*(y^*) + x_J^*(y^*) + 2Cy^* + 2C(y_0 - y^*)e^{(2C-\delta)t} > 0 \end{aligned}$$

provided $y_0 < y^*$, which is a natural assumption.

Suppose player i is not happy about the above optimal choice of the bloc and quits the bloc at time T. Then, the above optimal choice continues until $t = T$ and pollution accumulation reaches

$$y(T) = (y_0 - y^*)e^{(2C-\delta)T} + y^* \tag{9}$$

where T is to be determined and it depends on the choice of strategic space after the splitting.

The total payoff of player i just before the splitting is thus

$$\alpha W(T) = \alpha \left[\frac{(a_i^2 + a_J^1 - 2B^2)(1 - e^{-rT})}{2r} - 2BC \int_0^T e^{-rt} y(t) dt - (C^2 + b) \int_0^T e^{-rt} y^2(t) dt \right]. \tag{10}$$

It is straightforward that

$$\frac{dW(T)}{dT} = e^{-rT} \left[\frac{a_i^2 + a_J^2 - 2B^2}{2} - 2BCy(T) - (C^2 + b)y^2(T) \right] > 0 \tag{11}$$

if and only if

$$y(T) = (y_0 - y^*)e^{(2C-\delta)T} + y^* \in (0, \ \underline{y})$$

where

$$\underline{y} = \frac{-2BC + \sqrt{4B^2C^2 + 2(C^2 + b)(a_i^2 + a_J^2 - 2B^2)}}{2(C^2 + b)} (> 0). \tag{12}$$

Obviously, the analysis above provides an upper-bound condition for remaining in the bloc in terms of pollution:

Corollary 1 *Under the assumptions of Proposition 1, provided $\alpha > 0$:*

- *if the initial condition checks $y_0 > \underline{y}$, then $T = 0$;*
- *if the coalition potential long-run steady state checks $(y_0 <) y^* < \underline{y}$, then $T = +\infty$.*

The above corollary can be written in a more compact manner:

$$T \begin{cases} = 0, \ \text{if} \ \underline{y} < y_0, \\ \in (0, +\infty), \ \text{if} \ y_0 < \underline{y} < y^*, \\ = +\infty, \ \text{if} \ \underline{y} > y^*. \end{cases}$$

From the definition of $\underline{y}$ in (12), it is easy to see that $\lim_{a_i, a_J \to 0} \underline{y} = 0$, for all $b > 0$. By continuity when a_i, a_J are sufficiently small, and for $y_0 > 0$, it follows $\underline{y} < y_0$, thus, $T = 0$. In other words, when both bloc J and i have a low enough development level, the coalition hardly exists. Additionally, it can be shown[7] that for any a_i, a_J not both zeros, $\lim_{b \to 0} \underline{y} \geq +\infty > \lim_{b \to 0} y^* = \frac{1}{\delta}$. Thus, when the damage from pollution accumulation is low, as long as a_i and a_J are not both close to zero, the coalition remains optimal forever.

It's somehow easy to interpret the previous results. Given the payoff specifications, larger values for the technological parameter a give more utility for given pollution (this is the technological progress advantage) but this parameter multiplies linearly emissions, x. In contrast, larger b values yield more disutility from pollution but since this parameter multiplies the stock of pollution quadratically, participating into a coalition makes sense especially when b is small enough relative to the technological parameter. As a result, if there is no technological progress (at least for one player), then coalitions make no sense if $b > 0$ whatever $\alpha \neq 0$. Conversely, if b tends to zero, and technological progress is nonzero, then coalitions would last forever whatever $\alpha \neq 0$. These mechanisms will play an important role in all the cases we will scrutinize along the way.

To study the interior situation where splitting happens in finite time, we impose the following condition on the parameters:

Assumption 1 *The model parameters ensure that the following inequality holds:*

$$y_0 < \underline{y} < y^*.$$

Unfortunately, it's not possible to analytically explore how this condition relates to the deep parameters of the model given the expressions of $\underline{y}$, y^*, B and C. We shall fortunately

[7] It is easy to check that

$$\begin{aligned}
\underline{y} &= \frac{\sqrt{2(C^2+b)(a_i+a_J) - 4bB^2} - 2BC}{2(C^2+b)} \\
&\geq \left[\frac{\sqrt{C^2 + b - (4bC^2/(r+\delta-2C)^2)} - 2C^2}{2(C^2+b)} \right](a_i + a_J) \\
&\equiv S(b)(a_i + a_J)
\end{aligned}$$

and

$$y^* = \frac{r+\delta}{\delta(r+\delta) - 4b}(a_i + a_J).$$

Thus, one sufficient condition for $\underline{y} > y^*$ is $S(b) > y^*$, for any a_i and a_J not both zero. By l'Hopital's rule, $\lim_{b\to 0} S(b) = +\infty$.

obtain more interpretable expressions for the interior splitting conditions (and optimality) in the next section under the above assumption.[8]

3.2 Optimal strategies in the non-cooperative stage (after T)

Suppose player i exits the bloc at time T. For $t \geq T$, we consider only Markovian subgame perfect strategies, where the strategy is defined as the choice variable x_j for player $j = i, J$, which depends on time and the current state: $x_i(t) = x_i(t, y(t))$, for all y. Furthermore, given for $t \geq T$, the game is autonomous and defined over infinite time horizon, $[T, +\infty)$, thus we can directly study stationary Markovian perfect equilibria (MPE) via stationary Bellman value functions.

Define the Bellman Value function of player $j = i, J$ as $U_j(y)$, which must check the following HJB equation for $t \geq T$,

$$rU_j(y) = \max_{x_j} \left[a_j x_j - \frac{x_j^2}{2} - \frac{b\,y^2}{2} + U_j'(y)\ \ (x_i + x_J - \delta y) \right], \quad j = i, J.$$

From these HJB equations, Appendix A.1 demonstrates the following existence results of MPE.

Proposition 2 *Suppose that player i quits the bloc at time T, and that after the splitting both players i and J adopt Markovian strategies. Then there exists a stable affine Markovian subgame perfect Nash equilibrium*

$$(x_i^m, x_J^m) = (a_i + B^m + C^m y,\ a_J + B^m + C^m y)\,, \quad \forall y,$$

with coefficients

$$C^m = \frac{(r + 2\delta) - \sqrt{(r + 2\delta)^2 + 12b}}{6} (< 0), \quad B^m = \frac{(a_i + a_J)C^m}{r + \delta - 3C^m} (< 0).$$

For a given initial condition at T, the corresponding optimal state trajectory is

$$y^m(t) = (y(T) - \widehat{y^m})e^{(2C^m - \delta)(t-T)} + \widehat{y^m}, \quad \forall t \geq T,$$

where $\widehat{y^m} = \dfrac{a_i + a_J + 2B^m}{\delta - 2C^m} (> 0)$ is asymptotically stable long-run steady state.

[8] It's not difficult to see that Assumption 1 is not void through numerical exercises.

A direct corollary from the above two propositions is the following, where the detail proof is given in Step 3 of Appendix A.1.

Corollary 2 *Under the assumptions of Propositions 1 and 2, it follows*

$$y^* < \widehat{y^m}, \quad \forall r, \delta, b > 0.$$

In other words, player i may not be happy with the outcome from cooperation. However from the pollution accumulation point of view, cooperation still can do better than competition in terms of long-term pollution, which is a common outcome in the environmental agreements literature (see for example, Calvo and Rubio, 2013). Obviously, the decision to split cannot be entirely determined by the steady state pollution levels as we will see below.

We have so far characterized the optimal strategies for each stage of the game, and identified some useful properties along the way. We now move to the determination of the optimal splitting time (if it exists).

4 Optimal splitting time and its constitutional vs technological drivers

Under the above Markovian perfect Nash equilibrium, it is easy to check that social welfare in the second period of player i is

$$\begin{aligned} W_{i,II}^m &= \int_T^{+\infty} e^{-rt} \left(a_i x_i - \frac{x_i^2}{2} - \frac{by^2}{2} \right) dt \\ &= \frac{a_i^2 - (B^m)^2}{2} \int_T^{+\infty} e^{-rt} dt - B^m\, C^m \int_T^{+\infty} e^{-rt} y^m(t) dt - \frac{((C^m)^2 + b)}{2} \int_T^{+\infty} e^{-rt} (y^m)^2 dt, \end{aligned} \tag{13}$$

in which $y^m(t)$ depends also on the splitting time T. In order to assess how the splitting time affects player i's welfare, we compute $\frac{dW_{i,II}^m(y(T))}{dT}$ using (13):[9]

[9] See Appendix A.2 for the details

$$\frac{dW^m_{i,II}(y(T))}{dT} = \left\{[-rA^m_i + B^m(a_i + a_J + 2B)] + [B^m(2C - \delta - r) + (a_i + a_J + 2B)C^m]y(T) + [(2C - \delta) - r/2]C^m y^2(T)\right\} e^{-rT} = [\widehat{a} + \widehat{b}y(T) + \widehat{c}y^2(T)]e^{-rT} \begin{cases} < 0 \text{ for } 0 \leq y(T) < \overline{y}, \\ > 0 \text{ for } y(T) > \overline{y}, \end{cases} \tag{14}$$

with $\overline{y}$ positive root of the second degree polynomial $\frac{dW^m_{i,II}(y(T))}{dT} = 0$ [10] and $A^m_i = \dfrac{a_i^2}{2r} + \dfrac{(a_i + a_J)B^m}{r} + \dfrac{3(B^m)^2}{2r}$.

If $y(T) > \bar{y}$, then $\frac{dW^m_{i,II}(Y(T))}{dT} > 0$. Hence the later the splitting happens, if there is splitting at all, the higher the social welfare will be in the second period for player i. If this is the case, player i would postpone the splitting as late as possible. In other words, splitting will never happen when the pollution accumulation is at a sufficiently high level, such as $y(t) > \bar{y}$.

Recall that Assumption 1 states the condition under which the accumulation of pollution is increasing over time. Thus if we assume that players i and J are initially in a coalition and that $y_0 > \overline{y}$, then $y(t) > \overline{y}$ for all $t \geq 0$, and there will be no splitting. We conclude in the following corollary

Corollary 3 *Suppose Assumption 1 holds and that $y_0 > \overline{y}$, then splitting will never happen, that is, $T = +\infty$.*

Combining the above two corollaries, Corollaries 1 and 3, the relationship between $\overline{y}$ and $\underline{y}$ is not essential in determining the optimal splitting time T, as long as they both are located between y_0 and y^*, then there may be splitting in finite time. We study this issue in detail in the next sub-section.

[10] The positive root is given by
$$\overline{y} = \frac{-\widehat{b} - \sqrt{\widehat{b}^2 - 4\widehat{a}\widehat{c}}}{2\widehat{a}}$$
where $\widehat{a} = [-rA^m_i + B^m(a_i + a_J + 2B)] < 0$, $\widehat{c} = [(2C - \delta) - r/2]C^m > 0$ and $\widehat{b} = B^m(2C - \delta - r) + (a_i + a_J + 2B)C^m$.

4.1 Optimal finite splitting time

In order to focus on the situation where splitting can happen in finite time, Assumption 1 would be complemented by the following:

Assumption 2 *Suppose the initial condition and the parameter set check*

$$y_0 < \overline{y}.$$

Obviously, if $y^* < \overline{y}$, splitting may happen in finite time. If instead $y_0 < \overline{y} < y^*$, and if there is splitting, then splitting can only happen before the pollution accumulation reaches to the upper limit of splitting, $\overline{y}$. Otherwise, player i can not afford the damage cost from accumulated pollution and would rather stay with player J.

Substituting the first and second periods' welfare derivatives with respect to splitting time T, i.e., (11) and (14), into the first order condition $\alpha\frac{dW(T)}{dT} + \frac{dW^m_{i,II}}{dT} = 0$, it follows that the above first order condition is equivalent to:

$$\Lambda\,(y^m_T)^2 + \Sigma\, y^m_T + \Gamma = 0, \tag{15}$$

where $y^m_T = y^m(T)$ stands for the stock of pollution at the switching time and the coefficients are

$$\begin{cases} \Lambda = -\alpha\,(C^2 + b) + C^m\left(2C - \delta - \frac{r}{2}\right), \\ \Sigma = -2\alpha BC + B^m\,(2C - \delta - r) + C^m\,(a_i + a_J + 2B)\,, \\ \Gamma = \frac{1}{2}\alpha\,(a_i^2 + a_J^2 - 2B^2) - rA^m_i + B^m\,(a_i + a_J + 2B)\,. \end{cases} \tag{16}$$

The roots of (15), if exist, are given by

$$y^m_T = \frac{-\Sigma \pm \sqrt{\Sigma^2 - 4\Lambda\Gamma}}{2\Lambda}. \tag{17}$$

Obviously, the existence of real roots is granted if and only if $\Sigma^2 - 4\Lambda\Gamma \geq 0$. The last inequality condition is ensured by $\Lambda \leq 0$ and $\Gamma \geq 0$, which are equivalent to

$$\alpha \geq \frac{C^m\,(2C - \delta - r/2)}{C^2 + b} \equiv G(b) \tag{18}$$

and

$$\alpha \geq \frac{1+\left[\frac{3(C^m)^2}{(r+\delta-3C^m)^2}-\frac{4CC^m}{(r+\delta-3C^m)(r+\delta-2C)}\right]\left(\frac{a_J}{a_i}+1\right)^2}{\left(\frac{a_J}{a_i}\right)^2+1-\left[\frac{2C^2}{(r+\delta-2C)^2}\right]\left(\frac{a_J}{a_i}+1\right)^2} \equiv F\left(\frac{a_J}{a_i}, b\right). \tag{19}$$

In other words, finite time splitting $T^m \in (0, \infty)$ is guaranteed by the payoff share under coalition to be large enough for given pollution damage coefficient, b. It can be readily shown that $G(b)$ is increasing in b and $G(0) = \frac{1}{2}$ (see Appendix for all related computations). Consequently, the larger b, the larger the payoff share α needed to make finite time splitting possible. Moreover, as function $G(.)$ is increasing, α should be always bigger than one-half whatever b. As explained below Corollary 1 above, the larger b, the more player i is reluctant to join a coalition in a world with positive technological progress. Only when the payoff share α is large enough, would player i join the coalition to eventually leave it some time after.

Again, as for condition (18), finite time splitting is granted if the payoff share is large enough, although the involved lower bound in this case is different: in contrast to condition (18), the lower bound also depends on the technological gap, $\frac{a_J}{a_i}$. We shall examine the implications below.

It should be noted that the above condition is a first-order condition, we now move to the second-order one. Given that parameters Λ, Σ and Γ are independent of the switching time T, the second-order sufficient condition, $\alpha\frac{d^2W(T)}{dT^2} + \frac{d^2W^m_{i,II}}{dT^2} < 0$, holds if and only if

$$[2\Lambda y^m(T) + \Sigma] \ \frac{\partial y^m(T)}{\partial T} < 0.$$

Due to the assumption that pollution accumulation is increasing over time, that is, $\frac{\partial y^m(T)}{\partial T} > 0$ is always true for any T, then the second-order sufficient condition holds if and only if

$$2\Lambda y^m_T + \Sigma < 0. \tag{20}$$

The above second order condition is equivalent to

$$\pm\sqrt{\Sigma^2 - 4\Lambda\Gamma} < 0.$$

If $\Lambda < 0$, this implies that the only optimizer of $\alpha W(T) + W^m_{i,II}(T)$ is at

$$y^m_T = \frac{-\Sigma - \sqrt{\Sigma^2 - 4\Lambda\Gamma}}{2\Lambda}.$$

So the optimal finite splitting time, T, is unique. At the minute notice that combining Conditions (18) and (19) ensures the existence of a unique optimal finite splitting time T. We summarize this important result in the following proposition, the detailed proof is reported in Appendix A.3.

Proposition 3 *Let Assumptions 1 and 2 hold. Suppose player i quits the bloc at time T and after that player i and J adopt MPE given by Proposition 2. Suppose the sharing parameter α checks*

$$\max\left\{F\left(\frac{a_J}{a_i}, b\right),\ G(b)\right\} < \alpha < 1. \tag{21}$$

where functions $G(b)$ and $F\left(\frac{a_J}{a_i}, b\right)$ are defined in (18) and (19). Furthermore, suppose that the pollution quantity

$$y^m_T = \frac{-\Sigma - \sqrt{\Sigma^2 - 4\Lambda\Gamma}}{2\Lambda} \tag{22}$$

satisfies

$$y_0 < y^m_T < y^*,$$

where Λ, Σ, Γ are given by (16). Then player i optimally quits the bloc at a finite time T:

$$T = \frac{1}{2C - \delta} \ln\left(\frac{y^m_T - y^*}{y_0 - y^*}\right). \tag{23}$$

If Condition (21) fails, then it can happen that either $T = 0$ and splitting starts immediately or $T = +\infty$ and there is no splitting at all, or other possibilities which complement the more direct results of Corollary 1 on the absence of optimal finite time splitting. We shall pay more attention to the "corner" solution, $T = +\infty$ in the policy implications part of this section.

Note that Proposition 3 delivers an explicit solution for the optimal splitting time as none of the terms involved in (23) depends on T. It should be also noted that while the optimal splitting time, given by (23), does depend on the sharing parameter α through the term y^m_T, the pollution stock at the splitting time, its existence (in short, the second-order optimality conditions) does also depend on the latter. Indeed, Condition (21), which guarantees that the obtained finite separation time T is optimal, features the role of the sharing payoff parameter in the sustainability of the coalition at the optimum. Interestingly enough, Condition (21) yields that the three fundamental ingredients of our model do a priori matter in the genesis of the split and the duration of coalitions: the sharing parameter, the pollution damage parameter, b, and the technological gap, $\frac{a_J}{a_i}$, do all matter.

The economic interpretation of Condition (21) involves the technological gap: for given technological gap and pollution damage parameter, the payoff share under coalition is required to be large enough for player i to engage in a coalition and to stay in for a finite time. Again, the constitutional parameter α is key in the optimal institutional dynamics: it's key for the existence of an optimal finite time splitting, and it's also key for the duration of the coalition (through the level of pollution at the splitting time, y_T^m as explained above). We shall devote the next subsection to the latter point. In the meanwhile, we shall clarify the implications of Proposition 3 by specializing in two cases depending on the new parameter (with respect to conditions (18) and (19)), that's parameter $\frac{a_J}{a_i}$). A more general result is stated and proved in the Appendix.

We shall indeed isolate two cases here, the case of a lagged country i, $\frac{a_J}{a_i} > 1$, and a case where this country is advanced.

Corollary 4 *Under the assumptions of Proposition 3, provided $\frac{a_J}{a_i} > 1$, then splitting occurs optimally at time T if and only if $\alpha > G(b)$. In all these splitting cases, $\alpha > \frac{1}{2}$.*

We prove in the Appendix that under $\frac{a_J}{a_i} > 1$, we get $F\left(\frac{a_J}{a_i}, b\right) < G(b)$, implying that the existence and optimality condition degenerates into the simpler condition $\alpha > G(b)$. Since $\frac{1}{2} < G(b) < \frac{1}{\sqrt{3}}$, $\forall b \geq 0$, it follows that finite time splitting occurs for any value of the damage parameter under the conditions specified in Proposition 3 and specialized in the corollary above. A technologically lagged country may join the coalition for any value of b provided the reward, as captured by α, is large enough, and in any case, larger than $\frac{1}{2}$. Notice that this is true whatever the value of the technological gap, provided it's bigger than one. Suppose now, the country i is advanced, what would be the outcomes? We give below a simple illustrative case.

Corollary 5 *Let the assumptions of Proposition 3 be satisfied. Assume that*

$$\frac{a_J}{a_i} < \frac{\sqrt{3} - 2 + \sqrt{12 - 2\sqrt{3}}}{2 + \sqrt{3}} \approx 0.7, \tag{24}$$

then:

$$F\left(\frac{a_J}{a_i}, b\right) \geq G(b)$$

holds for all $b \geq 0$. *Therefore, (21) holds if and only if*

$$F\left(\frac{a_J}{a_i}, b\right) < \alpha.$$

In all these splitting cases, $\alpha > \frac{1}{2}$.

In the case of a more technologically advanced country i, the technological gap shows up is the existence and optimality Condition (21) contrary to the case of a lagged player i studied above. It's not surprising: as we said, a lagged country will always benefit from a coalition if the pollution damage is small enough, at least for a while if their payoff share under coalition is good enough. The tradeoffs are more involved if the country is more advanced than the other coalition members (on average). Suppose that the pollution damage parameter, b, then the benefits for player i to join a coalition are rather thin, and one may expect that the more this player is advanced, that's the smaller the technological gap, $\frac{a_J}{a_i}$, the larger the payoff share requested by player i to join the coalition for a while. This is exactly what we show in the Appendix: for b small enough, function $F(x, b)$ is decreasing in x. In contrast, when the pollution damage parameter, b, is large, the function $F(x, b)$ is increasing in x for small values of x. In other words, if the pollution damage is viewed as big enough, the more advanced the country, the lower the payoff share requested to join the coalition for a while. Our model is therefore able to finely generate all the possible institutional configurations depending on three key parameters, $(\alpha, b, \frac{a_J}{a_i})$.

4.2 Payoff sharing and the duration of coalitions

In this section, we clarify how the sharing strategy α affects the duration of a coalition. We know that the duration of the coalition is increasing in y_T^m, the pollution stock at the splitting time, according to equation (23). A critical point, already also outlined in the section above is that the pollution stock at the splitting time, y_T^m, is increasing in α if $y_T^m \leq \underline{y}$ and decreasing in α if $y_T^m > \underline{y}$, where $\underline{y}$ is defined in (12) and y_T^m in (22). We can go a step further, Appendix A.5 shows the following results.

Theorem 1 *Suppose there exists $\alpha_0 \in [0,1]$, such that, $y_T^m(\alpha_0)$ satisfies $0 \leq y_T^m(\alpha_0) < \underline{y}$, then $y_T^m(\alpha)$ is increasing in α in a neighborhood of α_0 in $[0,1]$. Similarly, if there exists a value α_0 such that $y_T^m(\alpha_0) > \underline{y}$, then $y_T^m(\alpha)$ is decreasing in α in a neighborhood of α_0 in $[0,1]$.*

To better grasp the importance of Theorem 1, let us link Assumption 1 and the last statement in Proposition 3. Assumption 1 delivers a condition under which a coalition may split in finite time. The first assumption of Theorem 1 states that if there exists a sharing parameter, α_0, such that the coalition can split in finite time, then increasing player i's payoff share raises the stock of pollution upon splitting y_T^m, leading to more durable coalitions (by Proposition 3). Therefore, when the stock of pollution generated by the coalition is relatively small, one obtains that the larger the payoff share of player i, the later splitting (and the larger the subsequent pollution stock at the splitting time). Recall that postponing T increases joint payoff in the first stage of the game, thus in the latter case, lengthening the coalition duration benefits both players i and J.

The opposite holds true. If the payoff share of player i is such that the stock of pollution y_T^m is relatively high, then the coalition is actually no longer beneficial[11]. In this case, splitting is accelerated. Here, an increase in player i's payoff share fastens the splitting process: if the stock of pollution is relatively large, then increasing the payoff share of player i would lead to an earlier breakdown of the coalition.

The above results are based on one particular sharing strategy α_0, which may be difficult to find. The following corollary extends the condition in the above theorem from one particular point into an interval, which is easier to find and apply. A detailed proof is given in Appendix A.6.

Corollary 6 *Suppose $y_T^m(\alpha)$ is real and nonnegative in a subinterval $(\alpha_1, \alpha_2) \subset [0,1]$. Then, either $y_T^m(\alpha) \leq \underline{y}$ in the entire subinterval (α_1, α_2), or $y_T^m(\alpha) \geq \underline{y}$ in the entire subinterval (α_1, α_2). In particular, if*

$$\max\left\{\frac{C^m(2C-\delta-r/2)}{C^2+b},\ \frac{2\left[rA_i^m - B^m(a_i+a_J+2B)\right]}{a_i^2+a_J^2-2B^2}\right\} < 1,$$

then either $y_T^m(\alpha) \leq \underline{y}$ or $y_T^m(\alpha) \geq \underline{y}$ for all α that satisfies (21).

[11] See the discussion after (11)

4.3 Policy insights

We shall now point at some potentially interesting policy insights one can extract from our theoretical analysis. Let's first mention that while our model may seem too stylized to tackle "real" splitting problems, it's not more stylized than the typical two-stage or repeated games devoted to this topic, surveyed in the Introduction: while it has some ingredients of differential games, it's essentially a two-stage game, the first stage being the coalition stage. The main difference with respect to the vast majority of two-stage games in this topic is the fact that the duration of the first stage is itself the direct result of an individual decision of a particular player. It's very hard to argue that this latter trait is unrealistic, this is exactly how the environmental agreements/protocols and other political coalitions have been unravelled.

Our analysis brings several interesting results along Sections 3, 4.1 and 4.2. Let's stress one of it: as clearly stated in Corollary 4 and 5, only "big" enough countries may under certain conditions pretend to split in the sense that optimal finite time splitting requires $\alpha > \frac{1}{2}$. Recall that the payoff share parameter is determined by the "constitution" of the coalition, reflecting in particular the relative historical, geographic, demographic and economic weight of the countries. If the constitution is also meant to guarantee no-splitting, two avenues can be taken within our framework. One is to counterbalance the impact of too large payoff share (in the sense of Corollary 4 and 5) by adding penalties to the constitutions, making sure that penalties are increasing enough in the payoff share to discourage splitting. The second (non-exclusive) solution is to limit by constitution the payoff share so that everlasting coalitions are the unique optimal institutional arrangement. In our theory, such an arrangement can be possible under the following conditions summarized in the following proposition.

Proposition 4 *Suppose α is chosen according to*

$$\alpha < \min\left\{G\left(b\right), F\left(\frac{a_J}{a_i}, b\right)\right\},$$

thus violating Condition (21). Then, the coalition optimally lasts forever provided the initial pollution, y_0 is large enough.

Technically speaking, Proposition 4 delivers conditions under which the "corner" solution, $T = \infty$, is optimal. As already announced above, the full analysis of the corner solutions, including the one singled out in Proposition 4, is given in the Appendix. With respect to the optimal finite time splitting case, not only Condition (21) is violated: the optimality of the everlasting coalition also requires the initial pollution to be large enough, which indeed also violates the second condition stated in Proposition 3 (that is, $y_0 < y_T^m < y^*$).[12]. Therefore, a constitution which limits the payoff share (to the upper bound identified in Proposition 4) depending on the pollution damage and the relative technological position of each country would, in our model, allows to block splitting if the level of pollution is threatening enough. Of course, if the pollution level is small, then coalitions are much less attractive, and would not even exist in this case (that's the corner solution $T = 0$ would arise in such a case).

5 Conclusion

In this paper, we have presented and explored an alternative analytical approach to the splitting coalitions problems. We use a parsimonious two-stage differential game setting where the duration of the first stage, the coalitional stage, depends on the will of a particular player to leave the coalition through an explicit timing variable. By specializing in a standard linear-quadratic environmental model augmented with a minimal constitutional setting for the coalition establishment, we are able to analytically extract several non-trival findings. While our setting is stylized (in particular, minimal constitution, constant technological gap and no-renegotiation), we do believe that our full analytical approach allows to enlighten in a neat and intuitive way the main challenges inherent in splitting processes.

Three key aspects drive the results obtained: the technological gap as an indicator of heterogeneity across countries (here technological), the constitution of the coalition (here captured by the payoff share under coalition) and the intensity of the public bad (here, the pollution damage coefficient). In a regional/international environment with low (enough) productivity, coalitions can hardly be started if the pollution damage parameter is large enough. In contrast, in the polar case where the latter is low (enough), coalitions start and optimally last for ever in a productive environment. In intermediate parametric cases, the outcomes are much trickier. If the "swinging" country is technologically lagged, then it will always benefit from a coalition for a while if their payoff share is large enough.

[12]In the Appendix, we show that y_0 large enough corresponds indeed to $y_0 > y_T^m$

In such case, only the pollution damage parameter is relevant in the determination of payoff share required to join the coalition. In the case of a more technologically advanced country, the technological gap also matters in the coalition formation: if the pollution damage parameter is small (Resp. big), the benefits for the advanced country to join the coalition are thin (Resp. significant), and the more this player is advanced, the larger (Resp. lower) the payoff share requested by this player to join the coalition.

Under other mild technical conditions, in particular on the initial level of pollution, all these intermediate parametric cases lead to optimal coalition formation with finite duration, that's with finite time splitting. A key characteristic of these finite-time-lived coalitions is the requirement of the payoff share accruing to the splitting country to be large enough (bigger than one-half in our two-player game). Incidentally, our two-stage differential game setting reaches the conclusion that splitting countries are precisely those which use to benefit the most from the coalition. Constraining the payoff share to be low by Constitution to prevent splitting may optimally work (that's lead to optimal everlasting coalitions) only provided initial pollution is high enough, which may cover the emergency cases we are witnessing nowadays.

A Appendix: Proof

A.1 Proof of Proposition 2 and Corollary 2

The proof is completed in three steps: step 1 demonstrates the existence of affine-linear Markovian Nash equilibrium; step 2 shows the stability and step 3 provides steady states comparison.

Step 1. Existence of Markovian Nash equilibrium

Define the Bellman Value function of player $j = i, J$ as $U_j(y)$, which must check the following HJB equation: for $t \geq T$,

$$rU_j(y) = \max_{x_j} \left[a_j x_j - \frac{x_j^2}{2} - \frac{b\, y^2}{2} + U_j'(y)\ \ (x_i + x_J - \delta y) \right], \ \ j = i, J.$$

Then the first order condition yields

$$x_j^m(t) = a_j + U_j'(y(t)). \tag{25}$$

Guess

$$U_j(y) = A_j + B_j y + \frac{C_j}{2} y^2, \quad \text{and} \quad j = i, J,$$

then

$$U_j'(y) = B_j + C_j y.$$

Substituting $x_i = a_i + B_i + C_i y$ and $x_J^m(t) = a_J + B_J + C_J y(t)$ into the HJB equations, comparing coefficients on both hand sides, it yields

$$\begin{cases} rA_i = \frac{(a_i+B_i)^2}{2} + (a_J + B_J)B_i, \\ (r + \delta - C_i - C_J)B_i = C_i(a_i + a_J + B_J), \\ (r + 2\delta)C_i = C_i^2 + 2C_i\, C_J - b, \end{cases} \tag{26}$$

and

$$\begin{cases} rA_J = \frac{(a_J+B_J)^2}{2} + (a_i + B_i)B_J, \\ (r + \delta - C_i - C_J)B_J = C_J(a_i + a_J + B_i), \\ (r + 2\delta)C_J = C_J^2 + 2C_i\, C_J - b, \end{cases} \tag{27}$$

Remark. More generally, if $b_i \neq b_J$, then the b in the last two equations in (26) and (27) should be b_i and b_J respectively.

Solving the above two systems of equations simultaneously, it follows that the only coefficients which yields valid Bellman value functions are

$$C_i = C_J = \frac{(r + 2\delta) - \sqrt{(r + 2\delta)^2 + 12b}}{6} \equiv C^m, \text{ and } B_i = B_J = \frac{(a_i + a_J)C^m}{r + \delta - 3C^m} \equiv B^m,$$

and

$$A_j^m = \frac{a_j^2}{2r} + \frac{(a_i + a_J)B^m}{r} + \frac{3(B^m)^2}{2r}, \quad j = i, J.$$

Thus the Markovian Nash equilibrium is given by

$$(x_i^m, x_J^m) = (a_i + U_i'(y),\ a_J U_J'(y)) = (a_i + B^m + C^m y,\ a_J + B^m + C^m y)\,, \quad \forall y.$$

Step 2 Stability

The stability is straightforward by substituting the above Markovian Nash equilibrium into the state equation, it yields

$$\dot{y} = (a_i + a_J + 2B^m) + (2C^m - \delta)y \;\; \forall t \geq T$$

with $y(T)$ coming from the first period cooperation and T unknown. The explicit solution is thus straightforward as given in the Proposition. Furthermore, it is easy to obtain the long-run steady state

$$y^m(t) = (y(T) - \widehat{y^m})e^{(2C^m-\delta)t} + \widehat{y^m}(>0).$$

Given $2C^m - \delta < 0$, for any $y(T)$, the trajectory asymptotically converges to this steady state.

Step 3. Proof of Corollary 2.

We can easily rewrite the two steady states of pollution as

$$y^* = \frac{a_i + a_J + 2B}{\delta - 2C} = (a_i + a_J)\,\frac{r+\delta}{(\delta - 2C)(r + \delta - 2C)}$$

and

$$\widehat{y^m} = \frac{a_i + a_J + 2B^m}{\delta - 2C^m} = (a_i + a_J)\,\frac{r+\delta - C^m}{(\delta - 2C^m)(r + \delta - 3C^m)}.$$

Therefore, in order to compare which steady state yields higher pollution, we only need to compare

$$h_1 \equiv \frac{r+\delta}{(\delta - 2C)(r + \delta - 2C)} = \frac{r+\delta}{\delta(r+\delta) - 2(r+2\delta)C + 4C^2}$$

and

$$h_2 \equiv \frac{r+\delta - C^m}{(\delta - 2C^m)(r + \delta - 3C^m)} = \frac{r+\delta - C^m}{\delta(r+\delta) - 2(r+2\delta)C^m + 6(C^m)^2 - \delta C^m}.$$

It is easy to see that

$$C < C^m < 0, \;\; \forall r, \delta, b > 0.$$

Thus,

$$-C > -C^m > 0, \quad C^2 > (C^m)^2 > 0,$$

and

$$-2C(r+2\delta)+4C^2 > -2C^m(r+2\delta)+4(C^m)^2.$$

So it is straightforward that

$$h_1 = \frac{r+\delta}{\delta(r+\delta)-2C(r+2\delta)+4C^2} < \frac{r+\delta}{\delta(r+\delta)-2C^m(r+2\delta)+4(C^m)^2}.$$

Furthermore, simple algebra yields that

$$\frac{r+\delta}{\delta(r+\delta)-2C^m(r+2\delta)+4(C^m)^2} < \frac{r+\delta-C^m}{\delta(r+\delta)-2(r+2\delta)C^m+6(C^m)^2-\delta C^m} = h_2.$$

Hence,

$$h_1 < h_2 \text{ and } y^* < \widehat{y^m}, \quad \forall r,\delta,b>0.$$

That completes the proof.

A.2 The first order condition

In this section, we obtain the derivative of second period's welfare with respect to time, (14), and the first order condition at the same time.

Suppose $rA_i^m > B^m(a_i + a_J + 2B)$ and Assumption 1 holds. The switching time T is given by the FOC

$$\alpha\frac{dW(T)}{dT} + \frac{dW_{i,II}^m}{dT} = 0,$$

provided the second order sufficient condition holds:

$$\alpha\frac{d^2W(T)}{dT^2} + \frac{d^2W_{i,II}^m(T)}{dT^2} < 0.$$

By definition,

$$\begin{aligned} W_{i,II}^m &= \int_T^{+\infty} e^{-rt}\left(a_i x_i - \frac{x_i^2}{2} - \frac{by^2}{2}\right)dt \\ &= \int_T^{+\infty} \frac{e^{-rt}}{2}\left(a_i^2 - (B^m)^2 - 2B^m C^m y(t) - \left[(C^m)^2 + b\right]y(t)^2\right)dt \end{aligned}$$

where $y(t)$ is a function of $y^m(t)$, which depends on T. Direct calculation yields:

$$\begin{aligned}\frac{dW_{i,II}^m}{dt} &= \frac{e^{-rT}}{2}\left[-a_i^2+(B^m)^2+2B^mC^my(T)-\left((C^m)^2+b\right)y(T)^2\right]\\ &\quad+\int_T^{+\infty}\frac{e^{-rt}}{2}\left[-2B^mC^m\frac{\partial y^m}{\partial T}-2\left((C^m)^2+b\right)y^m\frac{\partial y^m}{\partial T}\right]dt.\end{aligned}$$

In order to obtain explicit result, we try to get rid of the term $\frac{\partial y^m}{\partial T}$ in the above first order derivative.

Let $V_j^\sigma(y)$ be the value function of Player j in Mode σ when the value of the state variable is y, for $j=i,J$ and $\sigma=1,2$, where $\sigma=1$ represents the mode before splitting, and $\sigma=2$ after splitting. In Mode 1 the players have the unchangeable Markovian strategies

$$x_j=a_j+B+Cy \qquad \text{for } j=i,J.$$

In addition, Player i has the impulse control in Mode 1 to exit the bloc. In Mode 2, the players have the value functions $V_j^2(y)=U_j(y)$ for $j=i,J$. The optimal strategy of Player i's impulse control results in maximization of the value $V_i^1(y)$ for $y<y^m$, where y^m is the value of y when Player i breaks up with the bloc. Hence, we have

$$V_i^1(y^m)=V_i^2(y^m)\equiv U_i(y^m) \tag{28}$$

and $\partial_{y^m}V_i^1(y)=0$ for any $y<y^m$. Since players are in Mode 1, Player i has the instantaneous utility

$$\alpha\left[a_ix_i+a_Jx_J-\frac{x_i^2+x_J^2}{2}-by^2\right]=\alpha\left[a_i^2+a_J^2-B^2-2BCy-\left(C^2+b\right)y^2\right],$$

and the dynamics of y is given by

$$\dot{y}=x_i+x_J-\delta y\equiv a_i+a_J+2B+(2C-\delta)y.$$

The HJB equation for V_i^1 is

$$rV_i^1=\alpha\left[a_i^2+a_J^2-B^2-2BCy-\left(C^2+b\right)y^2\right]+\frac{dV_i^1}{dy}\left[a_i+a_J+2B+(2C-\delta)y\right] \text{ for } y<y^m$$

subject to the terminal condition (28). The solution can be written in the integral form

$$
\begin{aligned}
V_i^1(y) &= U_i(y^m) + \int_{y^m}^{y} \frac{dV_i^1(z)}{dy} dz \\
&= U_i(y^m) + \int_{y^m}^{y} \frac{rV_i^1(z) - \alpha\left[a_i^2 + a_J^2 - B^2 - 2BCz - (C^2+b)z^2\right]}{a_i + a_J + 2B + (2C-\delta)z} dz
\end{aligned}
$$

for $y < y^m$. Differentiating the both sides with respect to y^m and using the condition $\partial_{y^m} V_i^1(y) = 0$ we find

$$
0 = U_i'(y^m) - \frac{rV_i^1(y^m) - \alpha\left[a_i^2 + a_J^2 - B^2 - 2BCy^m - (C^2+b)(y^m)^2\right]}{a_i + a_J + 2B + (2C-\delta)y^m}.
$$

Since

$$
\begin{aligned}
V_i^1(y^m) &= U_i(y^m) = A_i^m + B^m y^m + \frac{C^m}{2}(y^m)^2 \qquad \text{and} \\
U_i'(y^m) &= B^m + C^m y^m,
\end{aligned}
$$

we obtain that the first order condition is equivalent to

$$
\begin{aligned}
(B^m + C^m y^m)\left[a_i + a_J + 2B + (2C-\delta)y^m\right] = r\left(A_i^m + B^m y^m + \frac{C^m}{2}(y^m)^2\right) \\
-\alpha\left[a_i^2 + a_J^2 - B^2 - 2BCy^m - \left(C^2+b\right)(y^m)^2\right],
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{dW_{i,II}^m}{dT} &= \left\{-r\left(A_i^m + B^m y(T) + \frac{C^m}{2} y(T)^2\right)\right. \\
&+ \left.(B^m + C^m y(T))\left(a_i + a_J + 2B + (2C-\delta)y(T)\right)\right\} e^{-rT}.
\end{aligned}
$$

Combining with (11), we obtain

$$
\alpha \frac{dW(T)}{dT} + \frac{dW_{i,II}^m}{dT} = e^{-rT}\left\{\Lambda y^m(T)^2 + \Sigma y^m(T) + \Gamma\right\} \tag{29}
$$

where the coefficients are given in (16).

That completes the proof.

A.3 Proof of Proposition 3

The first order condition (29) can be rewritten as the following second degree polynomial in $y(T)$:

$$\Lambda\, y^m(T)^2 + \Sigma\, y^m(T) + \Gamma = 0.$$

The roots, if they exist, they are given by

$$y^m(T) = \frac{-\Sigma \pm \sqrt{\Sigma^2 - 4\Lambda\Gamma}}{2\Lambda}. \tag{30}$$

Since parameters Λ, Σ and Γ are independent of switching time T, the second order sufficient condition holds if and only if

$$(2\Lambda y(T) + \Sigma)\;\; y'(T) < 0.$$

Given the assumption that pollution is always increasing over time, that is, $y'(T) > 0$, then the second order sufficient condition holds if and only if

$$2\Lambda y(T) + \Sigma < 0. \tag{31}$$

Combining the second order condition (31) and the explicit solution (17), it follows that

$$2\Lambda y(T) + \Sigma = \pm\sqrt{\Sigma^2 - 4\Lambda\Gamma} < 0$$

if and only if the negative sign is taken in the explicit solution (17). Taking into account that only positive pollution levels are feasible, then

$$y^m(T) = -\frac{\Sigma + \sqrt{\Sigma^2 - 4\Lambda\Gamma}}{2\Lambda} > 0,$$

which is true if $\Lambda < 0$, $\Gamma > 0$ (these two inequality implicitly guarantee the existence of real positive solution from FOC) and regardless the sign of Σ. Condition $\Lambda < 0$ is equivalent to

$$\alpha > \frac{C^m\,(2C - \delta - r/2)}{C^2 + b} \equiv G(b)$$

and $\Gamma > 0$ if and only if

$$\alpha > \frac{2\left[rA_i^m - B^m\left(a_i + a_J + 2B\right)\right]}{a_i^2 + a_J^2 - 2B^2} \equiv F(a_J, a_i, b).$$

To finish the proof, from the explicit solution,

$$y(T) = (y_0 - y^*)e^{(2C-\delta)T} + y^* = y^m$$

and rearranging terms, it yields that

$$T = \frac{1}{2C - \delta}\ln\left(\frac{y^m - y^*}{y_0 - y^*}\right).$$

Recall Assumption 1 guarantees that $y_0 < y(T) < y^*$, thus, $0 < \frac{y^m - y^*}{y_0 - y^*} < 1$ and

$$T \in (0, +\infty).$$

That completes the proof.

A.4 Profs of Corollaries 4 and 5.

We first prove

Lemma 1 *Let the assumptions of Proposition 3 be satisfied. Then the following properties hold.*

1. *For a_J and a_i that satisfy*

$$\frac{a_J}{a_i} \leq \frac{\sqrt{3} - 2 + \sqrt{12 - 2\sqrt{3}}}{2 + \sqrt{3}}, \tag{32}$$

   relation

$$F\left(\frac{a_J}{a_i}, b\right) \geq G\left(b\right)$$

   holds for all $b \geq 0$.

2. For a_J and a_i that satisfy

$$\frac{\sqrt{3}-2+\sqrt{12-2\sqrt{3}}}{2+\sqrt{3}} < \frac{a_J}{a_i} < 1, \tag{33}$$

there is $b^*(a_J/a_i)$ such that

$$F\left(\frac{a_J}{a_i}, b\right) > G(b) \qquad \text{if and only if } b < b^*\left(\frac{a_J}{a_i}\right).$$

Hence, (21) holds if

$$\begin{aligned} F\left(\frac{a_J}{a_i}, b\right) &< \alpha \qquad \text{for } b < b^*\left(\frac{a_J}{a_i}\right) \text{ and} \\ G(b) &< \alpha \qquad \text{for } b > b^*\left(\frac{a_J}{a_i}\right). \end{aligned}$$

3. For a_J and a_i that satisfy

$$\frac{a_J}{a_i} \geq 1 \tag{34}$$

relation

$$F\left(\frac{a_J}{a_i}, b\right) \leq G(b)$$

holds for all $b > 0$.

Proof. Part 1. We rewrite $F(a_J, a_i, b)$ as

$$F\left(\frac{a_J}{a_i}, b\right) = \frac{1+\left[\frac{3(C^m)^2}{(r+\delta-3C^m)^2} - \frac{4CC^m}{(r+\delta-3C^m)(r+\delta-2C)}\right]\left(\frac{a_J}{a_i}+1\right)^2}{\left(\frac{a_J}{a_1}\right)^2+1-\left[\frac{2C^2}{(r+\delta-2C)^2}\right]\left(\frac{a_J}{a_i}+1\right)^2}.$$

For shorter notation, denote $x = \frac{a_J}{a_i}$, $H = \frac{3(C^m)^2}{(r+\delta-3C^m)^2}$, $K = \frac{4CC^m}{(r+\delta-3C^m)(r+\delta-2C)}$ and $L = \frac{2C^2}{(r+\delta-2C)^2}$, then

$$F(x,b) = F\left(\frac{a_J}{a_i}, b\right) = \frac{1+(1+x)^2(H-K)}{x^2+1-(1+x)^2 L}, \tag{35}$$

with $H - K < 0$ and $L < 1/2$. Thus, the condition on α can be shorten as:

$$\max\{G(b), F(x,b)\} < \alpha < 1.$$

Straightforward algebra yields that $\forall r, \delta > 0$,

$$\lim_{b\to 0} G(b) = \frac{1}{2}, \quad \lim_{b\to +\infty} G(b) = \frac{1}{\sqrt{3}}, \quad \text{and} \quad \frac{dG(b)}{db} > 0, \quad \lim_{b\to 0} \frac{dG(b)}{db} = 0,$$

thus

$$G(b) \in \left(\frac{1}{2}, \frac{1}{\sqrt{3}}\right), \quad \forall b > 0.$$

Again straightforward, though cumbersome, algebra yield that

$$\lim_{b\to 0} F(x,b) = \frac{1}{x^2+1}, \quad \lim_{b\to +\infty} F(x,b) = \frac{1-(1+x)^2/3}{x^2+1-(x+1)^2/2}. \tag{36}$$

Furthermore,

$$\frac{\partial F(x,b)}{\partial b} > 0 \qquad \text{if } x < \frac{\sqrt{2}}{2}, \qquad \frac{\partial F(x,b)}{\partial b} < 0 \qquad \text{if } x > \sqrt{\frac{3}{5}}$$

for all $b > 0$ and there is $\hat{b} > 0$ such that

$$\frac{\partial F(x,b)}{\partial b} = \begin{cases} > 0, & \text{if } 0 < b < \hat{b}, \\ < 0, & \text{if } b > \hat{b}, \end{cases} \qquad \text{if } \frac{\sqrt{2}}{2} \leq x \leq \sqrt{\frac{3}{5}}. \tag{37}$$

Note that

$$\frac{1-(1+x)^2/3}{x^2+1-(x+1)^2/2} > \frac{1}{\sqrt{3}}$$

if and only if (32) holds with $x = a_J/a_i$. Therefore

$$F(x,b) > \frac{1}{\sqrt{3}} \geq G(b) \qquad \text{for any } b > 0$$

if (32) holds. This proves Part 1.

Part 2. Since $x = a_J/a_i$ satisfies the reversed inequality in (32),

$$\lim_{b\to\infty} F(x,b) = \frac{1-(1+x)^2/3}{x^2+1-(x+1)^2/2} < \frac{1}{\sqrt{3}}.$$

In addition

$$\lim_{b\to 0} F(x,b) = \frac{1}{x^2+1} > \frac{1}{2} = \lim_{b\to 0} G(b),$$

by the intermediate value theorem, there is a $b^*(x)$ such that

$$G(b) = F(x, b^*(x)).$$

We show that $b^*(x)$ is the only solution to the above equation and

$$\begin{cases} G(b) < F(x,b), & \text{if } \ 0 < b < b^*(x), \\ G(b) > F(x,b), & \text{if } \ b > b^*(x). \end{cases} \tag{38}$$

Since

$$\frac{1}{\sqrt{2}} < \frac{\sqrt{3} - 2 + \sqrt{12 - 2\sqrt{3}}}{2 + \sqrt{3}} < \sqrt{\frac{3}{5}},$$

$F(x, b)$ is either decreasing in b for all b or there is a $\hat{b} > 0$ such that (37) holds if

$$x > \frac{\sqrt{3} - 2 + \sqrt{12 - 2\sqrt{3}}}{2 + \sqrt{3}}.$$

In the former case, since $G(b)$ is increasing, it is obvious that $b^*(x)$ is the only solution and (38) holds. In the latter case, $F(x, b)$ is bell-shaped. Note that

$$\lim_{b \to 0} F(x, b) = \frac{1}{x^2 + 1} > \frac{5}{8} > \frac{1}{\sqrt{3}} = \lim_{b \to \infty} G(b) \qquad \text{if } x < \sqrt{\frac{3}{5}},$$

it follows that

$$F(x, b) \geq \lim_{b \to 0} F(x, b) > \lim_{b \to \infty} G(b) \qquad \text{if } b < \hat{b}.$$

Therefore, $b^*(x) > \hat{b}$. Since $F(x, b)$ is decreasing in b for $b > \hat{b}$, we again find $b^*(x)$ is the only solution and (38) holds.

Part 3. Since $x \geq 1$, it follows that

$$\lim_{b \to 0} F(x, b) = \frac{1}{x^2 + 1} \leq \frac{1}{2} = \lim_{b \to 0} G(b).$$

Furthermore, since $x > \sqrt{3/5}$, $F(x, b)$ is decreasing in b. Hence, since $G(b)$ is increasing in b, we find

$$F(x, b) \leq \lim_{b \to 0} F(x, b) \leq \lim_{b \to 0} G(b) \leq G(b).$$

This completes the proof.

Now, Corollary 4 follows directly from Part 3 of Lemma 1, and Corollary 5 follows directly from Part 1 of Lemma 1.

A.5 Proof of Theorem 1

We first show that $\underline{y}$ is a positive real number. Since $C < 0$, it follows that

$$|B| = \frac{(a_i + a_J)\,|C|}{r + \delta + 2\,|C|} < \frac{a_i + a_J}{2}.$$

Hence,

$$a_i^2 + a_J^2 - 2B^2 > a_i^2 + a_J^2 - \frac{(a_i + a_J)^2}{2} = \frac{1}{2}(a_i - a_J)^2 \geq 0.$$

Therefore the numerator in (12) is positive. This proves the assertion.

Note that Λ, Σ, Γ and y^m all depend on α. We use $\Lambda(\alpha)$, $\Sigma(\alpha)$, $\Gamma(\alpha)$, and $y^m(\alpha)$ to indicate the dependence.

By Proposition 3, $y^m(\alpha)$ is a solution to the quadratic equation

$$\Lambda(\alpha)(y^m)^2 + \Sigma(\alpha)\,y^m + \Gamma(\alpha) = 0.$$

Let

$$F(y,\alpha) = \Lambda(\alpha)\,y^2 + \Sigma(\alpha)\,y + \Gamma(\alpha)\,.$$

Then, $F(y^m(\alpha),\alpha) = 0$ and

$$\frac{dy^m(\alpha)}{d\alpha} = -\frac{F_\alpha(y^m(\alpha),\alpha)}{F_y(y^m(\alpha),\alpha)}. \tag{39}$$

By differentiation,

$$\begin{aligned}
F_\alpha(y^m(\alpha),\alpha) &= \Lambda_\alpha(\alpha)(y^m(\alpha))^2 + \Sigma_\alpha(\alpha)\,y^m(\alpha) + \Gamma_\alpha(\alpha) \\
&= -\left(C^2 + b\right)y^m(\alpha)^2 - 2BCy^m(\alpha) + \frac{1}{2}\left(a_i^2 + a_J^2 - 2B^2\right), \\
F_y(y^m(\alpha),\alpha) &= 2\Lambda(\alpha)\,y^m(\alpha) + \Sigma(\alpha)\,.
\end{aligned}$$

It is shown in the Proof of Proposition 3 that $2\Lambda(\alpha)\,y^m(\alpha) + \Sigma(\alpha) < 0$ (see (20)). Note that $\underline{y}$ is the only positive solution to the quadratic equation

$$-\left(C^2+b\right)y^2-2BCy+\frac{1}{2}\left(a_i^2+a_J^2-2B^2\right)=0.$$

Hence, since $y^m(\alpha_0)<\underline{y}$,

$$F_\alpha\left(y^m(\alpha_0),\alpha_0\right)=-\left(C^2+b\right)\left(y^m(\alpha_0)\right)^2-2BCy^m(\alpha_0)+\frac{1}{2}\left(a_i^2+a_J^2-2B^2\right)>0.$$

It follows that $dy^m/d\alpha>0$ at a_0. Hence $y^m(\alpha)$ is nondecreasing in α in a neighborhood of α_0.

This proof for the other case is similar. This completes the proof.

A.6 Proof of Corollary 6

Suppose there are points $\hat{\alpha},\tilde{\alpha}\in(\alpha_1,\alpha_2)$ such that $y^m(\hat{\alpha})<\underline{y}<y^m(\tilde{\alpha})$. Then $y^m(\alpha)$ is increasing in α at $\hat{\alpha}$ and it is decreasing in α at $\tilde{\alpha}$. It is not possible that $\tilde{\alpha}<\hat{\alpha}$ because otherwise there is an $\tilde{\alpha}<\bar{\alpha}<\hat{\alpha}$ such that $y^m(\bar{\alpha})$ is the minimum of $y^m(\alpha)$ between $\tilde{\alpha}$ and $\hat{\alpha}$. Thus $dy^m(\bar{\alpha})/d\alpha=0$.

However, by (39), $F_\alpha\left(y^m(\bar{\alpha}),\bar{\alpha}\right)=0$, which implies that $y^m(\bar{\alpha})=\underline{y}$. Therefore, $y^m(\bar{\alpha})=\underline{y}>y^m(\hat{\alpha})$, which is a contradiction. So it is necessary that $\hat{a}<\tilde{a}$. In this case there is $\bar{\alpha}$ such that $\hat{\alpha}<\bar{\alpha}<\tilde{\alpha}$ and $y^m(\bar{\alpha})$ is the maximum of $y^m(\alpha)$ between $\hat{\alpha}$ and $\tilde{\alpha}$. Hence, again $dy^m(\bar{\alpha})/d\alpha=0$ and we have $y^m(\bar{\alpha})=\underline{y}<y^m(\tilde{\alpha})$. This is again a contradiction. Therefore no such points $\hat{\alpha}$ and $\tilde{\alpha}$ exist.

This completes the proof.

A.7 Optimal splitting time: interior vs corner solutions

The optimal splitting time, T, satisfies the equation

$$\alpha\frac{dW(T)}{dT}+\frac{dW_{i,II}^m(T)}{dT}=0.$$

By (11) and (14), the left-hand side can be written as

$$e^{-rT}\left[\Lambda y^m(T)^2+\Sigma y^m(T)+\Gamma\right]$$

where Λ, Σ, and Γ are given by (16). We define

$$\eta(y) = \Lambda y^2 + \Sigma y + \Gamma.$$

In the case where $\Lambda \neq 0$, $\eta(y)$ has two roots

$$y_1^m = \frac{-\Sigma - \sqrt{\Sigma^2 - 4\Lambda\Gamma}}{2\Lambda}, \qquad y_2^m = \frac{-\Sigma + \sqrt{\Sigma^2 - 4\Lambda\Gamma}}{2\Lambda}$$

which are real if

$$\Sigma^2 \geq 4\Lambda\Gamma,$$

and in the case where $\Lambda = 0$, $\eta(y)$ has one root

$$y_0^m = -\Gamma/\Sigma$$

provided that $\Sigma \neq 0$.

Using $G(b)$ and $F(a_J/a_i, b)$ defined in (18) and (19),

$$\Lambda < 0 \Longleftrightarrow \alpha > G(b), \qquad \Gamma > 0 \Longleftrightarrow \alpha > F\left(\frac{a_J}{a_i}, b\right).$$

There are four possible cases.

Case 1:

$$\alpha > \max\{G(b), F(a_J/a_i, b)\}.$$

In this case, $\Lambda < 0$ and $\Gamma > 0$. So η has one positive root, y_1^m. Furthermore, $\eta(y) > 0$ if $y < y_1^m$ and $\eta(y) < 0$ if $y > y_1^m$. So if $y(0) < y_1^m$, coalition can be formed and lasts until $y(T) = y_1^m$, and if $y(0) \geq y_1^m$, coalition cannot be formed.

Case 2:

$$F(a_J/a_i, b) \leq \alpha \leq G(b).$$

In this case, $\Gamma \geq 0$ and $\Lambda \geq 0$. There are four subcases, (1) $\Lambda = 0$ and $\Sigma < 0$, (2) $\Lambda = 0$ and $\Sigma \geq 0$, (3) $\Lambda > 0$ and $\Sigma^2 \leq 4\Lambda\Gamma$, and (4) $\Lambda > 0$ and $\Sigma^2 > 4\Lambda\Gamma$.

In subcase (1), $\eta(y)$ is linear and has one positive root, y_0^m. Also, $\eta(y) > 0$ if $y < y_0^m$ and $\eta(y) < 0$ if $y > y_0^m$. So if $y(0) < y_0^m$ coalition lasts until $y(T) = y_0^m$, and if $y(0) \geq y_0^m$

coalition cannot be formed.

In subcases (2) and (3), $\eta(y) \geq 0$ for all $y > 0$. Therefore, coalition lasts forever.

In subcase (4), $\eta(y)$ has two positive roots y_1^m and y_2^m if $\Sigma < 0$. It is clear that $y_1^m < y_2^m$, and $\eta(y) > 0$ for $y < y_1^m$ or $y > y_2^m$, and $\eta(y) < 0$ for $y_1^m < y < y_2^m$. So if $y(0) < y_1^m$, coalition continues until pollution reaches y_1^m. If $y_1^m \leq y(0) \leq y_2^m$, coalition cannot be formed, and if $y(0) > y_2^m$, coalition lasts forever.

Case 3:

$$G(b) \leq \alpha \leq F(a_J/a_i, b).$$

In this case $\Lambda \leq 0$ and $\Gamma \leq 0$. There are four subcases, (1) $\Lambda = 0$ and $\Sigma \leq 0$, (2) $\Lambda = 0$ and $\Sigma > 0$, (3) $\Lambda < 0$ and $\Sigma^2 \leq 4\Lambda\Gamma$, and (4) $\Lambda < 0$ and $\Sigma^2 > 4\Lambda\Gamma$.

In subcases (1) and (3), $\eta(y)$ is nonpositive for all $y > 0$. So a coalition cannot be formed.

In subcase (2) $\eta(y) < 0$ if $y < y_0^m$ and $\eta(y) > 0$ if $y > y_0^m$. So if $y(0) < y_0^m$ a coalition cannot be formed, and if $y(0) > y_0^m$ then, the coalition lasts forever.

In subcase (4), both y_1^m and y_2^m are nonnegative, and $\eta(y) < 0$ for $y < y_1^m$ or $y > y_2^m$ and $\eta(y) > 0$ for $y_1^m < y < y_2^m$. So if $y(0) < y_1^m$ or $y(0) > y_2^m$, a coalition cannot be formed, and if $y_1^m \leq y(0) \leq y_2^m$, then the coalition continues until $y(T) = y_2^m$.

Case 4:

$$\alpha < \min\{G(b), F(a_J/a_i, b)\}.$$

In this case $\Gamma < 0$ and $\Lambda > 0$. So η has one positive root, y_2^m. Also, $\eta(y)$ changes from negative to positive as y passes through y_2^m. So if $y(0) < y_2^m$, a coalition cannot be formed, and if $y(0) \geq y_2^m$, the coalition lasts forever.

Note that Proposition 4 follows from this conclusion.

References

[1] Battaglini M. and B. Harstad (2016). Participation and duration of environmental agreements. *Journal of Political Economy*, 1, 160-204.

[2] Bertinelli L., C. Camacho and B. Zou (2014). Carbon capture and storage and transboundary pollution: A differential game approach. *European Journal of Operational Research*, 237, 721-728.

[3] Boucekkine, R., W. Ruan and B. Zou (2022). The irreversible pollution game. *DEM working paper*, $2022-3$, University of Luxembourg.

[4] Boucekkine R., J. Krawczyk and T. Vallée (2011). Environmental quality versus economic performance: a dynamic game approach. *Optimal Control Applications and Methods* 32, 29-46.

[5] Boucekkine R., Pommeret A. and F. Prieur (2013). Optimal regime switching and threshold effects. *Journal of Economic Dynamics and Control*, 37, 2979-2997.

[6] Boucekkine, R., C. Camacho and B. Zou (2020). Optimal switching from competition to cooperation: a preliminary exploration. Chapter 9 in: Dynamic Economic Problems with Regime Switches, edited by Vladimir M. Veliov, Josef L. Haunschmied , Raimund Kovacevic and Willi Semmler, Springer.

[7] Bréchet, T., F. Gerard and H. Tulkens (2011). Efficiency vs. stability in climate coalitions: A conceptual and computational appraisal. *The Energy Journal*, 32, 49-75.

[8] Bucher B. C. Carraro and I. Cersosimo (2002). Economic consequences of the US withdrawal from the Kyoto/Bonn Protocol. *Climate Policy*, 2 (4), 273-292.

[9] Calvo, E., and S. Rubio, S. (2013). Dynamic models of international environmental agreements: A differential game approach. *International Review of Environmental and Resource Economics*, 6, 289-339.

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

[11] Dockner E. and N. Van Long (1993). International pollution control: cooperative versus noncooperative strategies. *Journal of Environmental Economics and Management* 24, 13-29.

[12] Dockner E., S. Jorgensen, N. Van Long , and G. Sorger (2000). Differential Games in Economics and Management. Cambridge University Press.

[13] Dutta, P., and R. Radner (2009). A strategic analysis of global warming: Theory and some numbers. *Journal of Economic Behavior and Organization*, 71, 187-209.

[14] Harstad, B. (2012). Climate contracts: A game of emissions, investments, negotiations, and renegotiations. *Review of Economic Studies*, 79, 1527-1557.

[15] Hattori, T. (2001). Renegotiations in the post-agreement period: the future of the Kyoto protocol. *Interdisciplinary Environmental Review*, 3, 64-80.

[16] Hoel, M. (1993). Intertemporal properties of an international carbon tax. *Resource and Energy Economics*, 15, 51-70.

[17] Latorre M., Z. Olekseyuk and H. Yonezawa (2020). Trade and foreign direct investmentrelated impacts of Brexit. *The World Economy.* 43, 232.

[18] Moser E., A. Seidel and G. Feichtinger (2014). History-dependence in production-pollution-trade-off models: a multi-stage approach. *Annals of Operations Research* 222, 455-481

[19] Nong D. and M. Siriwardana (2018). Effects on the U.S. economy of its proposed withdrawal from the Paris Agreement: A quantitative assessment. *Energy*, 159, 621-629.

[20] Saglam C. (2011). Optimal pattern of technology adoptions under embodiment: A multistage optimal control approach. *Optimal Control Applications and Methods* 32, 574-586

[21] Sampson T. (2017). Brexit: The Economics of International Disintegration. *Journal of Economic Perspectives* 31 (4), 163184.

[22] Tahvonen, O. (1994). Carbon dioxide abatement as a differential game. *European Journal of Political Economy* 10, 685-705.

[23] Tomiyama K. (1985). Two-stage optimal control problems and optimality conditions. *Journal of Economic Dynamics and Control* 9, 317-337

[24] Tulkens, H. (1998). Cooperation vs. free riding in international environmental affairs: two ap- proaches, in: N. Hanley and H. Folmer (eds), "Game Theory and the Environment", Chapter 2, 30- 44, Edward Elgar, Cheltenham.

[25] Xepapadeas A. (1995). Induced technical change and international agreements under greenhouse warming. *Resource and Energy Economics*, 17(1), 1-23.

[26] Zampolli F. (2006). Optimal monetary policy in a regime switching economy: the response to abrupt shifts in exchange rate dynamics. *Journal of Economic Dynamics and Control* 30, 15271567.