How do coalitions break down? An alternative view*

Raouf Boucekkine, Carmen Camacho, Weihua Ruan, Benteng Zou

How do coalitions break down? An alternative view*

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

Abstract

We propose an alternative dynamic theory of coalition breakdown. Motivated by recent coalition splitting events through unilateral countries' withdrawals, we assume that: i) the payoff sharing rule within coalitions is not necessarily set according to any optimality and/or stability criterion, and, ii) players initially behave *as if* the coalition will last forever. If the sharing rule is non-negotiable or if renegotiation is very costly, compliance to these rules may become unbearable for a given member because the rule, being too rigid, would make exit preferable as time passes. We examine this endogenous exit problem in the case of time-invariant sharing rules. Assuming a Nash non-cooperative game after (potential) splitting where players play Markovian, we characterize the solutions of the endogenous exit problem in a linear-quadratic frame with endogenous splitting time. We find that splitting countries are precisely those which use to benefit the most from the coalition. Suitable sharing rules should be used to prevent coalition splitting. When initial pollution is high, all shares should be low enough and none of the players should detain a payoff share larger than 1/2. If initial pollution is small, we provide with an explicit interval for the sharing rule values preventing the collapse of the coalition. Finally, we demonstrate that the latter properties are qualitatively consistent with the optimal behaviors and equilibrium outcomes resulting from players anticipating the end of the coalition and acting accordingly.

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

*In memory of Ngo Van Long. Financial support from the French National Agency (grant ANR-17-EURE-0001) is gratefully acknowledged by Carmen Camacho.

†Corresponding author. Centre for Unframed Thinking (CUT), 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; Senior Fellow, Centre for Unframed Thinking (CUT), Rennes School of Business, France. E-mail: Wruan@pnw.edu

¶DEM, University of Luxembourg, Luxembourg. E-mail: benteng.zou@uni.lu.

1 Introduction

While the classical literature on coalitions had essentially addressed the question of coalition formation and stability (with a few exceptions though, see Bolton et al., 1996), numerous papers on coalition break-ups have been written in the last few years. This abundant literature is essentially motivated by the recent numerous withdrawals of countries from international organizations and agreements, some highly impactful. Beside some of the decisions taken by the Trump administration, which may seem "idiosyncratic"[1], the United Kingdom withdrawal from the European Union on January 31, 2020, or Canada withdrawal from Kyoto Protocol on December 13, 2011, 10 years after the US, are two of these striking break-up events which have attracted the attention of economists and political scientists. Just as a way of illustration let us mention papers investigating the impact of Brexit (Sampson, 2017; La Torre et al., 2020; the special issue of the Oxford Review of Economic Policy, vol 33, 2017; etc.), or the economic consequences of the U.S. withdrawal from the Kyoto Protocol and the Paris agreement (Bucher et al., 2002; Zhang et al., 2017; Nong and Siriwardana, 2018; ...). [2]

This paper is a methodological contribution to this rising literature. The traditional game-theoretical settings proposed to study the design of international agreements and the stability of coalitions are quite diverse: they range from cooperative to non-cooperative, from static to dynamic through repeated games or fully dynamic set-ups, and they often include interesting procedural ingredients, typically on enforceability of the agreements. Regarding the particular problem of coalition splitting, one finds two main conceptual settings. The first one is based on the theory of coalition stability, which is anchored in the cooperative games literature.[3] A second type of setting uses the traditional Nash non-cooperative theory with individual optimizing strategies (plus a Pareto-like criterion to evaluate efficiency). An essential part of this literature uses dynamic games. Ngo Van Long (see, for example, Van Long, 2010, for a survey) is one of the principal contributors to this line of research.

Our paper departs from the latter dynamic games literature in two major ways:

- First of all, the sharing rule within coalitions is not necessarily set according to any optimality and/or stability criterion.
- Second, players initially behave *as if* the coalition will last for ever.

[1] For example: on July 7, 2020, the Trump administration formally notified the United Nations that the U.S. was pulling out of the World Health Organization, which became effective as of July 6th, 2021.

[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.

[3] Two different frames are considered in this literature. The first looks at static or repeated games as explained by Tulkens (1998) and surveyed by Bréchet et al. (2011). The second deals instead with dynamic games and, after Petrosjan (1977), searches for sharing mechanisms that will ensure the coalition stability (see Zaccour, 2007, for more details).

At first glance, the two departures seem rather realistic. The coalitions are typically based on a number of, say, constitutional rules specifying the duties and benefits corresponding to each member of the coalition. Usually, the coalition may obviously entail large heterogeneities across members, in particular in multi-country coalitions, technological, demographic or geographic notably. It's unlikely that the constitutional rules at the dawn of the coalition can cope with all these discrepancies, and meet any kind of optimality in the sense of the criteria given above (for example, Nash bargaining or Shapley value). It's also quite reasonable that when engaging in essential (and somehow existential) collective initiative like a strategic political alliance or a long-term environmental agreement, no member will start playing against it, they will rather act optimally in accordance with the constitutional rules, *as if* the coalition will last till the end time agreed upon. If the constitutional rules are non-negotiable or if renegotiation is very costly, compliance to these rules may become unbearable as time passes, either because the rules, being too rigid, would make exit preferable (endogenous exit) or because an exogenous (symmetric or asymmetric) shock occurs undermining the political, economic or historical rational behind the coalition creation. We shall consider the endogenous exit problem in this paper, specializing in the case of time-invariant and non-negotiable constitutional rules. As we shall see, the problem remains largely nontrivial in this benchmark case.

While realistic and fitting a variety of situations in very different contexts, our assumptions entail two theoretically unpleasant features: predetermined (non-optimal) shares under the coalition setting and time inconsistency. As we will see clearly when solving the dynamic games posed, our alternative frame involves the absence of forward-looking behavior in the coalition stage, which enables a forward induction solution method. We study this case till its ultimate consequences, including policy implications (which does make perfect sense as this case is based on reasonably realistic assumptions). Nonetheless, we also provide with the solution to the standard forward-looking counterpart where players do anticipate the coalition breakdown from $t = 0$ and behave accordingly. Despite that the induced solution scheme is opposite to the counterpart in our alternative case and consists in backward induction, we show that the ultimate optimal splitting problems are analogous. Moreover the main policy implication (related to the sharing rule) is qualitatively the same. That's to say the equilibrium outcomes generated by our alternative frame is quite far from irrationality (if rationality corresponds to the pure forward-looking case).

In our theory, since we focus on coalition splitting, we specialize in the simple case where one single country can potentially break down its tie to the coalition. After the split, if any, a two-players non-cooperative game sets in: the splitting country and the remaining coalition treated as a single player. This mimics many of the recent coalition withdrawals occurrences mentioned above, and we believe the theoretical approach taken is deep enough to highlight some of the key determinants of coalition breakdowns as we will argue throughout the text. Another critical aspect of our theory is (non)-negotiability. Non-negotiable here means that the core principles establishing the initial

coalition (say, the sharing rules) cannot be redefined to strictly cope with the preferences of any single country. For instance, we have in mind the case of the U.S. splitting from the 2015 Paris agreement or the case of UK splitting from the EU. We do precisely formalize this problem by designing a tractable dynamic game-theoretic frame assuming time-invariance of the sharing rules under coalition.

In this paper, we specialize in the theoretical literature of environmental agreements to fix the ideas. Fundamental work has already been accomplished on these issues 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). Here, we take a different approach to better suit the question of coalition splitting as we have posed it just above. We consider a set of countries which are initially bound in a coalition, and whose aim it to maximize a given joint payoff subject to a public bad. The coalition is based on time-invariant (typically suboptimal) constitutional rules, namely the sharing quotas of the benefits and the costs of the coalition. Given all these ingredients, under which conditions a country initially belonging to this coalition may eventually optimally decide to split at a finite date, and **when**? What are the determinants (Constitution, technology,...etc) of splitting and of the duration of the coalition? Is it possible to identify time-invariant consititional rules which prevent the coalition breakdown?

In our framework, the splitting time is an explicit optimal control in the hands of any member of initial coalitions. The recent contribution of Colombo et al. (2022) is close to our setting in investigating partial cooperation in international environmental agreements. There, all players are identical and so is the share of each player in joint welfare (regardless of whether there is full or partial cooperation). In their setting, the coalition that is optimally set at the initial time will last forever. Our work is also related to the seminal contribution of Benchekroun et al (2006), who study the temporary natural resource cartels where the cartels' ending time is known for all players at the beginning of the game, while in our setting it is a decision of one player. More precisely, we assume that initially players (countries) agree to manage cooperatively the common stock of pollution. 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 do not include splitting costs (formalizing possible penalties paid by the splitting country) for simplicity. As it will be clear in the main text and in the Appendix, the algebraic developments needed without this additional ingredient are already huge. Note that a splitting fixed sunk cost can hardly change the qualitative results in terms of the sustainability of time-invariant constitutional rules.

If a country splits at time T, a non-cooperative game sets in between the country and the group of countries remaining in the coalition. Within a fully 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 where all coalition members play *as if* the coalition will last for ever. As argued above, if this assumption fails, then a different

solution setting via backward induction should be applied reflecting the forward-looking nature of splitting in such a case. We ultimately uncover the conditions under which splitting occurs at finite time. We also study the determinants of the coalition duration with particular attention to the role of technological vs constitutional heterogeneity across players. It's worth pointing out that the choice of strategies can be extended to non-Markovian ones. For example, open-loop Nash equilibria or heterogeneous Nash equilibria *à la* Zou (2016) may match better some situations. For example, after the U.S. withdrawal from the Paris agreement, the remaining coalition stayed committed to the initial decarbonization objectives. Nevertheless, the techniques developed in this paper are general enough to study different kinds of choices of strategic spaces.[4]

Technically, our analytical approach combines multistage optimal control tools with the typical techniques used to solve differential games. There exist an increasing number of papers using multistage optimal control to characterize optimal/ equilibrium regime transitions and the inherent optimal regime shift timings (Boucekkine et al., 2013; Moser et al., 2014; Saglam, 2011; Zampolli et al. 2016; etc.).[5] In contrast, much fewer papers merging multi-stage optimal control and dynamic games have come out.[6] We shall show how the latter avenue can also be taken safely in our paper.

Three key aspects drive the paper's 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 to these parsimonious specifications, we are able to provide with a full 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. We specially highlight the requirement that the payoff share accruing to the splitting country should be large enough in the latter case. Consistently, we prove that constraining the payoff share to be low enough 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 presents the general specification of our game-theoretical setting. Section 3 analyzes a specialized linear-quadratic version of the game, providing in particular the optimal players' strategies for given splitting times. Section 4 characterizes the existence of an optimal splitting time, discusses its drivers and delivers some policy insights. Section 5 addresses the case of an uncommited coalition where the players anticipate splitting and behave in a forward-looking way in this respect. Finally, Section 6 concludes.

[4] In Appendix B, we illustrate this point by showing how our setting can be readily adapted to heterogeneous strategies after the splitting.

[5] Commonly, all these studies rely on Tomiyama (1984). It is worth mentioning here that our stopping time problem differs from the one explored in the literature under stochastic setting (see Shiryaev, 2008, and Albrecht et al, 2010). The main difference derives from the fact that our stopping time is a contingent event, not following any random observations.

[6] An exception is for example Boucekkine et al. (2011).

2 The model

Suppose that at time 0 there is a given coalition of players, say a pro-environmental coalition managing a common stock of pollution, denoted y. Suppose that one of the coalition members, named player i, can potentially quit it at some future date T, where $0 \leq T \leq \infty$. The rest of the coalition is assumed to be roughly homogenous to the point that we can label it as a single player J. Both players i and J differ in a number of characteristics, technological and constitutional as mentioned in the Introduction. Our analysis will identify *a posteriori* the specific characteristics that lead player i to split.

Within the coalition, 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. The players' actions increase the level of the public bad, y, resulting in a drop in welfare, which corresponds to the pollution externality in environmental economics. In our model, we assume that at time 0, players play cooperatively until time T, when player i decides eventually to quit the coalition. 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 after T. Finally, we introduce a further simplification by assuming that actions $x_i, x_J \in [0, X] \subset [0, +\infty)$, while determining the level of players' utilities does also increase the level of CO_2 emissions by exactly $x_i + x_J$. The model can be then straightforwardly interpreted in a one-good economy: the good x is consumed and produced with a linear technology, and there is a one-to-one relationship between input and output, and between output and pollution emissions. In the end, player j can obtain utility directly from x_j, but she also suffers from pollution, since y brings a (partially external) pollution disutility.

Initially, the objective of the players in the coalition is therefore to maximize joint overall welfare or payoff, which is defined for everlasting coalitions 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 utility functions of players i and J, respectively, which are strictly increasing and concave; and $c_i(y), c_J(y)$ are their respective individual disutility due to pollution, which are strictly increasing and convex functions. Note that the objective function is simply the aggregate payoff of the two players. We shall discuss 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:

$$\dot{y}(t) = x_i(t) + x_J(t) - \delta y(t), \tag{2}$$

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

$y(0) = y_0 \geq 0$ is given.

We get now into 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. 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 coalition, 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. 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 rule in a pollution 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 nor the Shapley value: 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 each country's weight also depends on more complex characteristics like global and regional history, 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 latter to clearly discriminate between the constitutional aspects of the coalition and the more purely technological diversity. We consider this case as the natural benchmark to explore. Moreover, we assume that renegotiation (of α) is impossible or too costly, which is far from unrealistic if one recalls the political and constitutional foundations of international coalitions like those mentioned in the Introduction: often splitting may prove less costly than renegotiation.[7]

[7] One way to avoid that player i quits the coalition is to redefine α as the Shapley value, allow renegotiation and letting α be a function of the common resource, i.e. $\alpha = \alpha(y)$.

We now move to some preliminary technical considerations. If player i quits the coalition at time T, then she obtains a share α of overall welfare until time T. Accordingly 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:

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

where the initial condition $y(T)$ is determined (by continuity) from the outcomes of the first (coalition) period.

The optimal switching time for player i is defined as

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

Intuitively, a coalition between i and J established at time 0 can last over the period of time $[0, T]$ if the first term in (6) is non-decreasing in T. That is, the longer player i stays in, the higher is joint social welfare. Otherwise, if the first term in (6) is decreasing in T, then player i would exit immediately the coalition with J and T will be 0. Similarly, player i may consider to quit the coalition with J if 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 $T = +\infty$ and somehow very surprisingly, this a priori non-optimal (almost ad-hoc) coalition would be stable and last forever.

Obviously, the precise optimal choice of T relies on the game that is played after the splitting, or more precisely, on the strategy space after the splitting. As one can deduce from all the above, the optimal choice of T can be 0, ∞ or take any other finite value between 0 and ∞, depending on the parameter set. If it exists, the interior optimal switching time T is obtained by taking the first order condition of (6), that is, T is the solution to

$$\alpha \frac{dW(T)}{dT} + \frac{dW_{i,II}}{dT} = 0, \tag{7}$$

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 using a linear-quadratic model, we first study the situation in which the coalition lasts forever, that is, when $T = +\infty$. Then, we will analyze the conditions ensuring the existence of a unique interior solution for T, $0 < T < +\infty$. If splitting occurs in finite time, then applying the implicit function theorem to (7) shows that

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

That is, the larger the payoff share player i gets, the later she will quit the coalition. Similarly, the smaller the stake of player i in the coalition, the sooner she quits the coalition to potentially gain more freedom of choice.

It is worth mentioning that differently from most of the optimal switching literature, the players before and after time T are different in our setting. Indeed, before time T, there is a single player: the coalition. After T, there are two competing players. Thus special care should be taken when employing the usual necessary optimal switching conditions at T. These difficulties may come mainly from the choice of different strategic spaces after the coalition splits. We shall be more explicit in this respect below.

3 The linear-quadratic differential game with initially committed coalition members

We start with the case where coalition members act *as if* the coalition lasts forever. As it's traditional in differential games, we resort to linear-quadratic functional forms for analytical tractability (see Dockner and Van Long, 1993; Dockner et al, 2000; Bertinelli et al, 2014; etc). In this section, we focus on the strategies at each stage of the game for a given splitting time, Section 4 will address the optimal splitting time issues. Section 5 considers the case of un-commited coalition members as mentioned in the Introduction.

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

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

If x_j is the pollution emission that player j employs to produce the final consumption good, then a_j is here the efficiency parameter which converts pollution into the consumption good. Note that 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.[8]

[8]Consistently with the general conditions on the preferences given in Section 2, the LQ utility functions posited

The pollution damage functions are

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

The pollution damage is the same for both players independent of any individual characteristic and in particular, independent of the agent's development level. In the following, we assume that coefficients a_j, $j = i, J$, are sufficiently large, such that the utility functions are always positive and increasing in x_j and that the long-run steady state of pollution is positive.

3.1 The cooperative stage

Let us start by solving the problem of the ever lasting coalition. As motivated in the Introduction, each coalition member acts *as if* the coalition lasts forever, and will do so till splitting (if any) occurs. The joint payoff function is

$$\max_{x_i, x_J} W(\infty) = \int_0^{+\infty} e^{-rt} \left[a_i x_i + a_J x_J - \frac{x_i^2 + x_J^2}{2} - by^2 \right] dt, \tag{8}$$

subject to the following state equation (2).

We can readily summarize the main results of the optimization problem faced by any coalition member acting ***as if* the coalition is everlasting**, in the following proposition.

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

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

where

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

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

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

where y^* is the asymptotically stable long-run steady state 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

are required to be increasing in the control domains. This amounts to having the controls x_k in the intervals $[0 \;\; a_k]$, $k \in \{i, J\}$. These conditions are checked for the optimal and equilibrium solutions computed hereafter.

levels, a_i and a_J. A higher technology level, which translates into higher consumption, leads to a higher level of long-run pollution. 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 is obviously a function of y^* and 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, aggregate consumption is always positive along the optimal trajectory path, that is,

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

$\forall y$ and $\forall r, \delta, a_i, a_J, b > 0$, provided $y_0 < y^*$, which is a natural assumption. If there exists a unique finite solution to (7), then using the Proposition above, the pollution stock would reach the value $y(T)$ given by

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

Accordingly, 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

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

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

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

- *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 (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 any $y_0 > 0$, it follows that $\underline{y} < y_0$, thus according to Corollary 1 it must necessarily be that $T = 0$. In other words, when both players in the initial coalition have a low enough development level, the coalition hardly exists. It is intuitive to see why. Both players here are heavy polluters, in the sense that they do not extract much consumption from the pollutant. None of them can make a remarkable effort to reduce pollution and alleviate the damage from the common bad. Since player i does not perceive the gains of staying in the coalition, she will exit immediately. Notice that here the motive for quitting the coalition is not free-riding in a strict sense since player i also takes into account J's welfare when they remain in the coalition. Here the coalition does not last because player i does no perceive any advantage for neither of them.

Additionally, it can be shown that for any a_i, a_J not both zero, $\lim_{b \to 0} \underline{y} \geq +\infty > \lim_{b \to 0} y^* = \dfrac{a_i + a_J}{\delta}$.[9] When pollution damage is low and a_i and a_J are not both close to zero, the coalition will remain together forever. Intuitively, when pollution damage is negligible, then one would have expected T to be 0 since there is no incentive to stay in the coalition. But, what we prove here is just the opposite. The reason for this seemingly contradicting result comes from the fact that when b is close to zero, both cooperation in the coalition and competition bring nearly identical welfare to player i. Indeed, note that the limit case shows $\lim_{b \to 0} x_j^* = a_j = \lim_{b \to 0} x_j^m$ for both $j = i, J$. Therefore, when damage from pollution is low and if the coalition is already established, then the coalition will last forever. However, if there was no coalition at time 0, then there is no incentive to form one either.

Let us add some further comments on our results. A given level of pollution stock provides on

[9] 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, $\lim_{b \to 0} S(b) = +\infty$.

the one hand with utility via consumption, but on the other hand, it also generates disutility. The positive effect on welfare is linear on emissions while the disutility generated by the stock of pollution is quadratic. Hence, participating in a coalition makes sense especially when b is small enough relative to the technological parameters a_i and a_J. 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 study in the following sections.

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

Assumption 1 *The model parameters ensure that the following inequalities hold:*

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

Unfortunately, it's not possible to explore analytically how this condition relates to the deep parameters of the model given the expressions of $\underline{y}$, y^*, B and C. We shall fortunately obtain interpretable expressions for the interior splitting conditions (and optimality) in the next section under the above assumption.

Remark 1 *Though it is not easy to clearly see if the parameter domains in which Assumption 1 holds are non-empty, it is not difficult to check it through numerical exercises. For example, let $r = 0.015$ and $\delta = 0.0005$. Considering that the damage parameter b is comparable with δ, we take $b \in [0.0001, 0.0002]$. Furthermore, given that a_i and a_J appear in Assumption 1 in the form of $a_i + a_J$, we set $a_i + a_J \in [0.3, 0.4]$. It can be shown that with the above parameters values, $y^* \in [19.7, 32]$ and $\underline{y} < y^*$. Obviously, y_0 can be chosen such that Assumption 1 holds. Nevertheless, as mentioned in the last footnote, for any $a_i + a_J > 0$, $\lim_{b \to 0} \underline{y} = +\infty > y^* = \frac{a_i + a_J}{\delta}$. Thus, with a sufficiently small damage parameter b, Assumption 1 fails to hold. It is also not so difficult to see that any combination of the parameters such that the numerator in $\underline{y}$ is negative leads to $\underline{y} < 0 < y_0$, and thus violates Assumption 1.*

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

Suppose player i quits the coalition at time T, and that after that both players play Markovian. Consider Markovian subgame perfect strategies: the strategies are such that the choice variables x_j for player $j = i, J$, depend upon time and the current state: $x_i(t) = x_i(t, y(t))$, for all y. Since the game is autonomous, we can directly study the stationary Markovian perfect equilibrium (MPE)

via the stationary HJB equations. If we denote the value functions of player $j = i, J$ as $U_j(y)$, they must check check the following HJB equations for $t \geq T$

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

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

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

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

with coefficients

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

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

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

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

The following corollary can be then obtained.

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

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

The proof is detailed at Step 3 of Appendix A.1. Even if the coalition breaks down at T because the option of staying in the coalition does not provide player i with higher welfare, the coalition does better in terms of pollution. This a standard outcome in the environmental agreements literature. Obviously, the decision to split can hardly be in general determined by the steady state pollution criterion, especially if players go Markovian after the split like in our case. This will be crystal clear in the next section devoted to the determination of the optimal splitting time.

4 Optimal splitting time and its drivers with initially committed coalition members

Under the above Markovian perfect Nash equilibrium, it is easy to obtain player i's welfare in the second period

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

where $y^m(t)$ also depends 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):[10]

$$\begin{aligned} \frac{dW_{i,II}^m(y(T))}{dT} = \; & \{[-rA_i^m + B^m(a_i+a_J+2B)] + [B^m(2C-\delta-r)+(a_i+a_J+2B)C^m]y(T) \\ & + [(2C-\delta)-r/2]C^m y^2(T)\}\, e^{-rT} = [\widehat{a}+\widehat{b}y(T)+\widehat{c}y^2(T)]e^{-rT} \\ & \begin{cases} <0 \;\text{ for }\; 0 \le y(T) < \overline{y}, \\ >0 \;\text{ for }\; y(T) > \overline{y}, \end{cases} \end{aligned} \tag{14}$$

with $\overline{y}$ the positive root of the second degree polynomial $\frac{dW_{i,II}^m(y(T))}{dT} = 0$ [11] and $A_i^m = \frac{a_i^2}{2r} + \frac{(a_i+a_J)B^m}{r} + \frac{3(B^m)^2}{2r}$.

If $y(T) > \bar{y}$, then $\frac{dW_{i,II}^m(Y(T))}{dT} > 0$. Hence the later the splitting happens, if there is splitting at all, the higher player i's welfare in the second period. If this is the case, then player i would postpone the splitting as much as possible. In other words, splitting never happens when the stock of pollution is high enough, namely if $y(t) > \overline{y}$.

Recall that Assumption 1 explicitly states the precise condition under which the stock 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 $t \ge 0$, and there will be no splitting. We conclude in the following corollary

[10] See Appendix A.2 for the details.

[11] The positive root is given by

$$\overline{y} = \frac{-\widehat{b} - \sqrt{\widehat{b}^2 - 4\widehat{a}\widehat{c}}}{2\widehat{a}}$$

where $\widehat{a} = [-rA_i^m + B^m(a_i+a_J+2B)] < 0$, $\widehat{c} = [(2C-\delta)-r/2]C^m > 0$ and $\widehat{b} = B^m(2C-\delta-r)+(a_i+a_J+2B)C^m$.

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

4.1 Optimal finite splitting time

In order to focus on the situation where splitting can happen in finite time, we must complete Assumption 1 with the following (given the properties established just above):

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

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

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

Substituting the first and second periods' welfare derivatives with respect to 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 first-order optimality 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 in (15) are

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

The roots of (15), if they 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

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

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, there exists a finite splitting time $T^m \in (0, \infty)$ if the payoff share is large enough, for given b. It can be readily shown that $G(b)$ is increasing in b and that $G(0) = \frac{1}{2}$ (see the Appendix for all related computations). Consequently, the larger b, the larger the 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, the larger b, the more reluctant player i is to stay in the coalition. Only when the payoff share α is large enough, would player i remain in the coalition (even if she will eventually leave it at some later time).

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 conditions result from the first-order condition, and we now move to the analysis of the second-order condition. 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

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

Since the pollution stock is increasing over time, we also have that $\frac{\partial y^m(T)}{\partial T} > 0$ 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 $\Lambda < 0$, then the unique 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, and its detailed proof is reported in Appendix A.3.

Proposition 3 *Let Assumptions 1 and 2 hold. Suppose player i quits the coalition at time T, and after that players 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*

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

satisfies

$$y_0 < y_T^m < y^*,$$

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

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

If Condition (21) fails to hold, then it can happen that either $T = 0$ and splitting is immediate, or $T = +\infty$ and there is no splitting at all. 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 also be noted that both the splitting time and the stock of pollution depend on the sharing parameter. Indeed, the optimal splitting time, given by (23), depends on α through y_T^m; and so do the level of the pollution stock at T and the conditions that ensure its existence (in short, the second-order optimality conditions). Interestingly enough, Condition (21) shows that the three fundamental ingredients of our model do matter in the duration of coalitions: the sharing parameter, the pollution damage parameter, b, and the technological gap, $\frac{a_J}{a_i}$.

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. Meanwhile, we shall clarify the implications of Proposition 3 by specializing in two cases depending on the ratio $\frac{a_J}{a_i}$:[12] first, the case of a technologically lagged country i, $\frac{a_J}{a_i} > 1$, and then a case where this country is more advanced.

Corollary 4 *Under the assumptions of Proposition 3, and provided* $\frac{a_J}{a_i} > 1$ *and* $\alpha > \frac{1}{2}$*, player* i *optimally quits the coalition at time* T *if and only if* $\alpha > G(b)$*.*

A technologically lagged country may remain in 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

[12] A more general result is stated and proved in Lemma 1 in the Appendix.

the value of the technological gap, provided it's bigger than one. Suppose now that country i is more advanced than J, what would the outcome be? We provide below a simple illustrative case.

Corollary 5 *Let the assumptions of Proposition 3 be satisfied and let* $\alpha > \frac{1}{2}$. *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*

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

Condition (21) is the single most important requirement in Proposition 3 since it provides a range for α for the player to remain in the coalition, even temporarily. The condition states that player i needs to retain a sufficiently large share of the total payoff to stay. In the case of a technologically advanced country i, the technological gap shows up in the existence and optimality Condition (21) contrary to the case of the lagged player i studied in Corollary 4. This is hardly surprising: a lagged country will always benefit from a coalition if the pollution damage is small enough, at least for a while, if her payoff share under coalition is good enough. However, as Corollary 5 shows, 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, is small, then the benefits for player i to remain in the coalition are rather thin. In this case one may expect that the more advanced player i's technology, that's the smaller the technological gap $\frac{a_J}{a_i}$, the larger the payoff share requested by player i to remain in the coalition. The contrary also holds true, if the pollution damage is large enough: in such a case, the more advanced the country, the lower the payoff share requested to remain in the coalition.[13] Proposition 3 and its corollaries show that our model is indeed able to generate finely 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

This section clarifies how the sharing parameter α affects the duration of a coalition. We already 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 is that the pollution stock at the splitting time, y_T^m, is

[13] We refer the reader to Appendix A.4 for a complete description and more general results.

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 2 *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 2, 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 in turn to a more durable coalition (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 the joint payoff in the first stage of the game, thus lengthening the coalition duration benefits both players i and J.

The opposite also 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 (for player i)[14] and an increase in α fastens the splitting process.

The above results are based on the existence of one particular sharing strategy α_0, which may be difficult to find. The following corollary extends the conditions in Theorem 2 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).

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

[14] See the discussion after (11).

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. There are two main differences with respect to the vast majority of dynamic or two-stage games in this topic. First, the initial coalition is defined by a profit sharing scheme that is not necessarily an equilibrium of any sort, so that the coalition is not necessarily stable. And second, the duration of the first stage is itself the direct result of the 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 them: as clearly stated in Corollaries 4 and 5, only "big" enough countries may under certain conditions quit the coalition. More precisely, optimal finite time splitting requires $\alpha > \frac{1}{2}$. Recall that the payoff share 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 Corollaries 4 and 5) by adding penalties to the constitution, 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 of all individual players, which guarantees 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*

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

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. With respect to the optimal finite time splitting case, not only Condition (21) is violated: the stability 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, $y_0 < y_T^m < y^*$).[15] 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, allow for coalition splitting if the level of pollution is threatening enough. Of course, if the pollution level is small, then coalitions are much less attractive, and they would not even exist. That is, the corner solution $T = 0$ would arise in such a case.

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

Finally, one can show that there is an interval of α on which the coalition lasts forever, without assuming any lower bound on initial pollution. Instead, we require a lower bound for the share α. Let

$$\underline{\alpha} = \frac{C^m \left(2C - \delta - r/2\right)}{C^2 + b}$$

and let $\hat{\alpha}$ satisfies the linear equation

$$\Lambda\left(\hat{\alpha}\right)\left(y^*\right)^2 + \Sigma\left(\hat{\alpha}\right) y^* + \Gamma\left(\hat{\alpha}\right) = 0$$

where y^* is defined by (22). Furthermore, let $\tilde{\alpha}$ be the first root greater than $\underline{\alpha}$ such that

$$\left(\Sigma\left(\tilde{\alpha}\right)\right)^2 - 4\Lambda\left(\tilde{\alpha}\right)\Gamma\left(\tilde{\alpha}\right) = 0.$$

If there is no such $\tilde{\alpha}$, we let $\tilde{\alpha} = +\infty$. Then, it can be shown that

$$0 < \underline{\alpha} < \min\left\{1, \hat{\alpha}, \tilde{\alpha}\right\}$$

and

Proposition 5 *Suppose $y_0 < y^*$. Then $T = \infty$ if α satisfies*

$$\underline{\alpha} < \alpha \leq \min\left\{1, \hat{\alpha}, \tilde{\alpha}\right\}. \tag{25}$$

The proof is given in Appendix A.8.

5 The linear-quadratic game with an initially uncommitted coalition member

In this section we consider that player i does not behave *as if* the coalition would last forever. That is, the player is forward-looking in what concerns splitting and inherent timing. To unburden the presentation and get the comparison with the "committed" case at glance, we shall focus on the identification of the coalition break-up point denoted in this section y_T^*, which is indeed the counterpart of y_T^m given by (22) in Subsection 4.1. Similarity between the two expressions will readily show up and confirm our claim in the Introduction.

Without the commitment assumption, the optimal controls $\left(x_i^*\left(t\right), x_J^*\left(t\right)\right)$ at the cooperative stage are not given by Proposition 1 because the joint value function at this stage is not a quadratic function of y. As a consequence the coalition breaks up at a different point. The next Proposition gives expression of y_T^*.

Proposition 6 *Let*

$$\begin{aligned}\lambda &= -\alpha\left(4\left(C^m\right)^2+b\right)+C^m\left(4C^m-\delta-\tfrac{r}{2}\right)\\ \sigma &= -8\alpha B^m C^m+B^m\left(4C^m-\delta-r\right)+C^m\left(a_i+a_J+4B^m\right)\\ \gamma &= \tfrac{\alpha}{2}\left[a_i^2+a_J^2-8\left(B^m\right)^2\right]-rA_i^m+B^m\left(a_i+a_J+4B^m\right)\end{aligned} \tag{26}$$

where B^m, C^m are given in Proposition 2, and A_i^m is given in (31) with $j=i$. Suppose that all players jointly optimize the utility

$$\int_0^T e^{-rt}\left[a_i x_i+a_J x_J-\frac{x_i^2+x_J^2}{2}-by^2\right]dt$$

in the coalition stage until the splitting time, T, and that the joint value function $W(y)$ for the coalition is differentiable. Then, the breaks up point $y_T^*=y(T)$ satisfies the equation

$$y_T^*=\frac{-\sigma-\sqrt{\sigma^2-4\lambda\gamma}}{2\lambda} \tag{27}$$

provided the right-hand side of (27) falls on the interval $(0,\infty)$.

The detailed proof is given in Appendix C. Comparing (26) and (27) with (16) and (22), we see that y_T^m becomes y_T^* if B and C in (16) are changed to $2B^m$ and $2C^m$, respectively. As a consequence, a similar sequence of results and implications can be derived with the same constructive approach as in the "committed" case, starting with a counterpart to the main Proposition 3. By construction and given the similarity of the algebra, the "uncommitted" and "commited" cases will deliver similar qualitative results. In particular, for coalitions to break down, the share α has to be large enough. Of course, the threshold values for α are no longer the same but the policy implications remain the same.[16]

6 Conclusion

In this paper, we have presented an alternative view of the coalition breakdown problem. Motivated by the most recent related events (in particular, unilateral withdrawal of several countries from institutional or environmental agreements), we have built up an alternative framework in which the sharing rules under the initial coalitions may not be necessarily optimal, and in which, because the coalitions are initially viewed as essential or existential, members act *as if* they will stay for ever. We have formulated and solved the corresponding, endogenous splitting problem assuming the renegotiation of the initial coalition consitutional rules are impossible or too costly.

[16] Comparison of the coalition durations in the two cases is not possible analytically. Our numerical exercises show however that the coalition survival is higher in the "committed" case, which is far from surprising.

We have solved the specific two-stage differential game induced by our alternative theory, and derived several highly nontrivial results on the technological/constitutional characteristics of the splitting coalitions and on the way the sharing rules could be set in order to prevent coalition breakdown. We have also solved the counterpart dynamic game where the players do anticipate splitting from $t = 0$ and act accordingly in a forward-looking manner. Despite the induced solution scheme is opposite to our alternative theory, the ultimate optimal splitting problems are shown to be analogous as well as the main policy (qualitative) implications. That's, while alternative, our framework is far from generating crazy outcomes. Of course, our analysis is preliminary in that it is for example restricted to time-invariant coalition payoff sharing rules. While our methodological approach can be still applied, analytical tractability is likely to be less conclusive than in our current benchmark.

A Appendix

A.1 Proof of Proposition 2 and Corollary 2

The proof is completed in three steps: step 1 demonstrates the existence of affine-linear Marvkovian 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

$$x_j^m(t) = a_j + U_j'(y(t)). \tag{28}$$

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

$$\begin{cases} rA_i = \frac{(a_i+B_i)^2}{2} + (a_J + B_J)B_i, \\ (r+\delta - C_i - C_J)B_i = C_i(a_i + a_J + B_J), \\ (r+2\delta)C_i = C_i^2 + 2C_i\, C_J - b, \end{cases} \tag{29}$$

and

$$\begin{cases} rA_J = \frac{(a_J+B_J)^2}{2} + (a_i + B_i)B_J, \\ (r+\delta - C_i - C_J)B_J = C_J(a_i + a_J + B_i), \\ (r+2\delta)C_J = C_J^2 + 2C_i\, C_J - b, \end{cases} \tag{30}$$

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

Solving the above two group equations system 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

$$A_j^m = \frac{a_j^2}{2r} + \frac{(a_i + a_J)B^m}{r} + \frac{3(B^m)^2}{2r}, \quad j = i, J. \tag{31}$$

Thus the Markovian Nash equilibrium is given by

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

Step 2 Stability

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

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

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 that 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 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,

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

That completes the proof.

A.2 The first order condition

In this section, we obtained 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 must be 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:

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

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 coalition. 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 coalition. Hence, we have

$$V_i^1(y^m) = V_i^2(y^m) \equiv U_i(y^m) \tag{32}$$

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

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

and the equation of dynamics is

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

The HJB equation for V_i^1 is

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

subject to the terminal condition (32). The solution can be written in the integral form

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

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

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

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

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

Combining with (11), we obtain

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

where the coefficients are given in (16).

That completes the proof.

A.3 Proof of Proposition 3

In the last subsection, the above first order condition can be rewritten as the following second degree polynomial in term of $y(T)$:

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

The roots, if they exist, are given by

$$y^{m}(T) = \frac{-\Sigma \pm \sqrt{\Sigma^{2} - 4\Lambda\Gamma}}{2\Lambda}. \tag{33}$$

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

$$2\Lambda y(T) + \Sigma < 0. \tag{34}$$

Combining the second order condition (34) 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 level is possible, 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 a real

positive solution from FOC) and regardless the sign of Σ. Condition $\Lambda < 0$ is equivalent to

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

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,

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

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 satisfies (24), 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 satisfies*

$$\frac{\sqrt{3} - 2 + \sqrt{12 - 2\sqrt{3}}}{2 + \sqrt{3}} < \frac{a_J}{a_i} < 1, \tag{35}$$

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

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

Hence, (21) holds if

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

3. *For* a_J *and* a_i *that satisfies*

$$\frac{a_J}{a_i} \geq 1 \tag{36}$$

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

$$F(x, b) = F\left(\frac{a_J}{a_i}, b\right) = \frac{1 + (1+x)^2(H-K)}{x^2 + 1 - (1+x)^2 L}, \tag{37}$$

with $H - K < 0$ and $L < 1/2$. Thus, the condition on α can be shortened 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 yields that

$$\lim_{b\to 0} F(x,b) = \frac{1}{x^2+1}, \quad \lim_{b\to+\infty} F(x,b) = \frac{1-(1+x)^2/3}{x^2+1-(x+1)^2/2}. \tag{38}$$

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

$$\frac{\partial F(x,b)}{\partial b} = \begin{cases} > 0, & \text{if } 0 < b < \hat{b}, \\ < 0, & \text{if } b > \hat{b}, \end{cases} \qquad \text{if } \frac{\sqrt{2}}{2} \le x \le \sqrt{\frac{3}{5}}. \tag{39}$$

Note that

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

if and only if (24) holds with $x = a_J/a_i$. Therefore

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

if (24) holds. This proves Part 1.

Part 2. Since $x = a_J/a_i$ satisfies the reversed inequality in (24),

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

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

$$\begin{cases} G(b) < F(x,b), & \text{if } 0 < b < b^*(x), \\ G(b) > F(x,b), & \text{if } b > b^*(x). \end{cases} \tag{40}$$

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 (39) 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 (40) 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 (40) 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 2

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

$$\frac{dy^m(\alpha)}{d\alpha} = -\frac{F_\alpha(y^m(\alpha),\alpha)}{F_y(y^m(\alpha),\alpha)}. \tag{41}$$

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

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

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 would exist an $\bar{\alpha}$, $\tilde{\alpha} < \bar{\alpha} < \hat{\alpha}$ 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 (41), $F_\alpha(y^m(\bar{\alpha}),\bar{\alpha}) = 0$, which implies that $y^m(\bar{\alpha}) = \underline{y}$. Therefore, $y^m(\bar{\alpha}) = \underline{y} > y^m(\hat{a})$, which is a contradiction. So it is necessary that $\hat{\alpha} < \tilde{\alpha}$. 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:

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

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:

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

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), $\eta\left(y\right)$ is linear and has one positive root, y_0^m. Also, $\eta\left(y\right) > 0$ if $y < y_0^m$ and $\eta\left(y\right) < 0$ if $y > y_0^m$. So if $y\left(0\right) < y_0^m$, the coalition lasts until $y\left(T\right) = y_0^m$, and if $y\left(0\right) \geq y_0^m$, a coalition cannot be formed.

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

In subcase (4), $\eta\left(y\right)$ 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 $\eta\left(y\right) > 0$ for $y < y_1^m$ or $y > y_2^m$, and $\eta\left(y\right) < 0$ for $y_1^m < y < y_2^m$. So if $y\left(0\right) < y_1^m$, coalition continues until pollution reaches y_1^m. If $y_1^m \leq y\left(0\right) \leq y_2^m$, a coalition cannot be formed, and if $y\left(0\right) > y_2^m$, the coalition lasts forever.

Case 3:

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

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), $\eta\left(y\right)$ is nonpositive for all $y > 0$. So a coalition cannot be formed.

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

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

Case 4:

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

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

Note that Proposition 4 follows from this conclusion.

A.8 Proof of Proposition 5

By computation,

$$\underline{\alpha} = \frac{r + 2\delta + \sqrt{(r+2\delta)^2 + 16b}}{2\left[r + 2\delta + \sqrt{(r+2\delta)^2 + 12b}\right]}. \tag{42}$$

It can be shown that the right-hand side is increasing in b. Hence

$$\frac{1}{2} < \underline{\alpha} < \frac{1}{\sqrt{3}} < 1.$$

We show that $\Sigma(\underline{\alpha}) > 0$. This is equivalent to showing that

$$B^m(2C - \delta - r) + C^m(a_i + a_J + 2B) > 2\hat{\alpha}BC. \tag{43}$$

By computation

$$\begin{aligned}
B^m(2C - \delta - r) &= \frac{2b(a_i + a_J)\left[r + \sqrt{(r+2\delta)^2 + 16b}\right]}{\left[r + \sqrt{(r+2\delta)^2 + 12b}\right]\left[r + 2\delta + \sqrt{(r+2\delta)^2 + 12b}\right]}, \\
C^m(a_i + a_J + 2B) &= \frac{-4b(a_i + a_J)(r+\delta)}{\left[r + \sqrt{(r+2\delta)^2 + 16b}\right]\left[r + 2\delta + \sqrt{(r+2\delta)^2 + 12b}\right]}, \\
BC &= \frac{32b^2(a_i + a_J)}{\left[r + \sqrt{(r+2\delta)^2 + 16b}\right]\left[r + 2\delta + \sqrt{(r+2\delta)^2 + 16b}\right]^2}.
\end{aligned}$$

Substituting the above and (42) into (43) and cancel common factors, the inequality is equivalent to

$$\frac{\left(r + \sqrt{(r+2\delta)^2 + 16b}\right)^2}{r + \sqrt{(r+2\delta)^2 + 12b}} - \frac{16b}{r + 2\delta + \sqrt{(r+2\delta)^2 + 16b}} > 2(r+\delta). \tag{44}$$

Note that

$$-\frac{16b}{r + 2\delta + \sqrt{(r+2\delta)^2 + 16b}} = r + 2\delta - \sqrt{(r+2\delta)^2 + 16b}.$$

The left-hand side of (44) can be written as

$$\frac{\left(r + \sqrt{(r+2\delta)^2 + 16b}\right)^2 + \left[r + 2\delta - \sqrt{(r+2\delta)^2 + 16b}\right]\left[r + \sqrt{(r+2\delta)^2 + 12b}\right]}{r + \sqrt{(r+2\delta)^2 + 12b}}. \tag{45}$$

Since

$$r+2\delta-\sqrt{(r+2\delta)^2+16b}<0, \qquad \sqrt{(r+2\delta)^2+12b}<\sqrt{(r+2\delta)^2+16b},$$

it follows that

$$\begin{aligned} & \left[r+2\delta-\sqrt{(r+2\delta)^2+16b}\right]\left[r+\sqrt{(r+2\delta)^2+12b}\right] \\ > & \left[r+2\delta-\sqrt{(r+2\delta)^2+16b}\right]\left[r+\sqrt{(r+2\delta)^2+16b}\right] \\ = & \left[r^2-(r+2\delta)^2-16b\right]+2\delta\left[r+\sqrt{(r+2\delta)^2+16b}\right]. \end{aligned}$$

Thus, the numerator of the quotient in (45) is greater than

$$\begin{aligned} & r^2+(r+2\delta)^2+16b+2r\sqrt{(r+2\delta)^2+16b}+\left[r^2-(r+2\delta)^2-16b\right]+2\delta\sqrt{(r+2\delta)^2+16b} \\ = & 2r^2+2(r+\delta)\sqrt{(r+2\delta)^2+16b}+2\delta r=2(r+\delta)\left[r+\sqrt{(r+2\delta)^2+16b}\right] \\ > & 2(r+\delta)\left[r+\sqrt{(r+2\delta)^2+12b}\right]. \end{aligned}$$

As a result, the quotient in (45) is greater than $2(r+\delta)$. This proves (44), which is equivalent to $\Sigma(\underline{\alpha})>0$.

Since $\Sigma(\underline{\alpha})>0$, $\Lambda(\underline{\alpha})=0$ and $\Lambda(\alpha)<0$ for $\alpha>\underline{\alpha}$, it follows from (22) that $y_T^m(\alpha)$ exists for $\alpha>\underline{\alpha}$ and is near $\underline{\alpha}$. In addition,

$$\lim_{\alpha\to\underline{\alpha}^+} y_T^m(\alpha)=\infty.$$

Hence, for $\alpha>\underline{\alpha}$ and is near $\underline{\alpha}$, $y_T^m(\alpha)>\underline{y}$. By Theorem 2 and Corollary 6, $y_T^m(\alpha)$ is decreasing in α for all $\alpha>\underline{\alpha}$ at which $y_T^m(\alpha)$ exists. As α increases, $y_T^m(\alpha)$ continues to exists until either α reaches $\tilde{\alpha}$ if $\tilde{\alpha}$ is finite, or α reaches $\hat{\alpha}$, or α reaches 1, whichever arrives earlier. If $\tilde{\alpha}\leq\min\{\hat{\alpha},1\}$, then for any $\alpha\in(\underline{\alpha},\tilde{\alpha})$

$$y_T^m(\alpha)>y^*. \tag{46}$$

By Proposition 3, $T=\infty$. If $\hat{\alpha}\leq\min\{\tilde{\alpha},1\}$, then (46) holds for $\alpha\in(\underline{\alpha},\hat{\alpha})$. Hence, $T=\infty$. Finally, if $1\leq\min\{\hat{\alpha},\tilde{\alpha}\}$, (46) holds for $\alpha\in(\underline{\alpha},1)$. So, $T=\infty$.

This completes the proof.

B Heterogeneous strategies

In this section, we illustrate one "out of equilibrium" outcome in which one player re-optimizes its behaviour after the fall of a coalition but the remaining members don't. The motivation draws

from the facts that Canada withdrew from Kyoto Protocol on December 13, 2011, the U.S. ceased its participation in the 2015 Paris Agreement on climate change mitigation in 2017 and the United Kingdom withdrew from the European Union on January 31, 2020. In all three examples, the remaining coalition did not change its strategy: neither did so the CO_2 emission targets in the Kyoto or Paris agreements, nor the trading rules within the EU.

More precisely, suppose player J stays with her original commitment to the coalition after T^i, $x_J^*(t)$. Thus, the differential game is reduced after T^i to a standard optimal control problem for player i:

$$\max_{x_i} W_{II}^i \equiv \int_{T^i}^{+\infty} e^{-rt} \left(a_i x_i - \frac{x_i^2}{2} - \frac{b\, y^2}{2} \right) dt,$$

subject to

$$\dot{y} = x_i + x_J^* - \delta y, \;\; \forall t \geq T^i,$$

with $y(T^i) = (y_0 - y^*)e^{(2C-\delta)T^i} + y^*$ and $x_J^*(y) = a_J + B + Cy$ given. The system is still autonomous and it is defined over an infinite time horizon. The same calculation as above yields the following:

Proposition 7 *Suppose that player i quits the coalition at time T^i, and that player J keeps her initial commitment. Then for any $t \geq T^i$, the optimal Markovian strategy of player i is*

$$x_i^i(y) = a_i + B^i + C^i y, \quad \forall y. \tag{47}$$

Furthermore, given the initial condition $y(T^i)$, the corresponding state variable $y^i(t)$ is given by

$$y^i(t) = \left[y(T) - \widehat{y}^i \right] e^{(C^i + C - \delta)(t - T^i)} + \widehat{y}^i \;\; \forall t \geq T^i,$$

where $\widehat{y}^i$ is the asymptotically stable long-run steady state and it is given by

$$\widehat{y}^i = \frac{a_i + a_J + B + B^i}{\delta - C - C^i},$$

and parameters

$$C^i = \frac{-2(C-\delta) - \sqrt{4(C-\delta)^2 + 4b(1-r)}}{2(1-r)} (< 0),$$

$$B^i = \frac{(a_i + a_J + B)C^i}{r + \delta - C - C^i} (< 0).$$

Obviously, since player J does not update her choice after the collapse of the coalition, the pair (x_J^*, x_i^i) may not be a Nash equilibrium given that x_J^* may not be the optimal response from player J's point of view, after the collapse of the coalition. Nonetheless, as discussed in Section 1, this is

one possible choice among others.

Straightforwardly, the separation time T^i can be obtained via the same technique as the one of Proposition 3 with similar, though different, parameter conditions.

The detailed calculations of the separation time T^i, as well as the comparison between T^i and the Markovian splitting time, T, together with their respective impacts on welfare can be found in Boucekkine et al (2023).

C Proof of Proposition 6

To derive (26) and (27), we use dynamic programming to obtain the equation

$$rW(y) = \max_{x_i, x_J} \left\{ a_i x_i + a_J x_J - \frac{x_i^2 + x_J^2}{2} - by^2 + W'(y)\left[x_i + x_J - \delta y\right] \right\}. \tag{48}$$

It is straighforward that the maximizers x_i^* and x_J^* of the right-hand side take the form

$$x_j^*(y) = a_j + W'(y) \qquad \text{for } j = i, J. \tag{49}$$

If the breaking up point $y_T^* \in (0, \infty)$ exists, W also satisfies the transition condition

$$W(y_T^*) = U_i(y_T^*) + U_J(y_T^*).$$

Since W, U_i and U_J are all differentiable, we also have

$$W'(y_T^*) = U_i'(y_T^*) + U_j'(y_T^*). \tag{50}$$

To find the optimal breaking up point for Player i, we let $W(y; y^*)$ to denote the solution of (48)-(49) for $y \in (0, y^*)$ and the boundary condition

$$W(y^*; y^*) = U_i(y^*) + U_J(y^*). \tag{51}$$

Consider a time t before breaking up when the stock of pollution is y. Let t^* be the time when the breaking up occurs. Then $t^* > t$. By dynamic programming,

$$W(y; y^*) = \int_t^{t^*} e^{-r(\tau - t)} g\left(y(\tau), x_i^*(y(\tau)), x_J^*(y(\tau))\right) d\tau + e^{-r(t^* - t)} \left[U_i(y^*) + U_J(y^*)\right] \tag{52}$$

where

$$g(y, x_i, x_J) = a_i x_i + a_J x_J - \frac{x_i^2 + x_J^2}{2} - by^2 \tag{53}$$

and $x_i^*(y)$ and $x_J^*(y)$ are given by (49). On the other hand, the total discounted benefit that Player i receives for the time interval (t,∞) is

$$\alpha\int_t^{t^*} e^{-r(\tau-t)}\left[a_i x_i^*(\tau)+a_J x_J^*(\tau)-\frac{x_i^*(\tau)^2+x_J^*(\tau)^2}{2}-by(\tau)^2\right]+e^{-r(t^*-t)}U_i(y^*).$$

We let $W_i(y;y^*)$ to denote this quantity. Using (52) we find

$$W_i(y;y^*)=\alpha W(y;y^*)+e^{-r(t^*-t)}\left[(1-\alpha)U_i(y^*)-\alpha U_J(y^*)\right]. \tag{54}$$

Clearly, if y_T^* is the optimal breaking up point for Player i, and if $0<y_T^*<\infty$, then the first and second order conditions

$$\frac{\partial W_i}{\partial y^*}(y;y_T^*)=0,\qquad \frac{\partial^2 W_i}{\partial y^{*2}}(y;y_T^*)\le 0 \tag{55}$$

both hold, since W_i is smooth.

We find $\partial W_i/\partial y^*$ as follows. Differentiating both sides of (54) with respect to t^*, we obtain

$$\begin{array}{r}\frac{\partial W_i}{\partial y^*}\frac{dy^*}{dt^*}=\alpha\frac{\partial W}{\partial y^*}\frac{dy^*}{dt^*}-re^{-r(t^*-t)}\left[(1-\alpha)U_i(y^*)-\alpha U_J(y^*)\right]\\ +e^{-r(t^*-t)}\left[(1-\alpha)U_i'(y^*)-\alpha U_J'(y^*)\right]\frac{dy^*}{dt^*}.\end{array} \tag{56}$$

Also, by differentiating both sides of (52) with respect to t^*, we find

$$\begin{array}{r}\frac{\partial W}{\partial y^*}\frac{dy^*}{dt^*}=e^{-r(t^*-t)}\left\{g(y^*,x_i^*(y^*),x_J^*(y^*))-r\left[U_i(y^*)+U_J(y^*)\right]\right\}\\ +e^{-r(t^*-t)}\left[U_i'(y^*)+U_J'(y^*)\right]\frac{dy^*}{dt^*}\end{array} \tag{57}$$

where dy^*/dt^* follows from the system of dynamic equations

$$\frac{dy^*}{dt^*}=x_i^*(y^*)+x_J^*(y^*)-\delta y^*. \tag{58}$$

We denote the right-hand side as $f(y^*)$. At $y^*=y_T^*$, the first order condition in (55) and the above two equations lead to

$$\alpha g(y_T^*)-rU_i(y_T^*)+U_i'(y_T^*)f(y_T^*)=0 \tag{59}$$

where $g(y^*)$ denotes $g(y^*,x_i^*(y^*),x_J^*(y^*))$. Note that U_i and U_J are quadratic functions

$$U_j(y)=A_j^m+B^m y+\frac{C^m}{2}y^2, \tag{60}$$

By (49) and (50),

$$x_j^*(y_T^*)=a_j+W'(y_T^*)=a_j+U_i'(y_T^*)+U_J'(y_T^*)\qquad \text{for } j=i,J \tag{61}$$

are linear functions of y_T^*. Hence, $f(y_T^*)$ is a linear function of y_T^* and $g(y_T^*)$ is a quadratic function

of y_T^*. This implies that equation (59) is quadratic in y_T^*. Thus we can write the equation in the form

$$\lambda \left(y_T^*\right)^2 + \sigma y_T^* + \gamma = 0,$$

and find the coefficients λ, σ, and γ using (60) and (61). The result is (26). This leads to

$$y_T^* = \frac{-\sigma \pm \sqrt{\sigma^2 - 4\lambda\gamma}}{2\lambda}. \tag{62}$$

It remains to show that the sign in front of the square root is negative. For this purpose we use the second order condition in (55). Using (58) we write (56) and (57) in the form

$$\frac{\partial W_i}{\partial y^*} f = \alpha \frac{\partial W}{\partial y^*} f - r e^{-r(t^*-t)} \left[(1-\alpha) U_i - \alpha U_J\right] + e^{-r(t^*-t)} \left[(1-\alpha) U_i' - \alpha U_J'\right] f. \tag{63}$$

and

$$\frac{\partial W}{\partial y^*} f = e^{-r(t^*-t)} \left\{g - r\left[U_i + U_J\right]\right\} + e^{-r(t^*-t)} \left[U_i' + U_J'\right] f, \tag{64}$$

respectively. Differentiating the two sides of (63) and (64) with respect to t^*, and using (55), we obtain

$$\begin{aligned} 0 \quad \geq \quad & \alpha \left[\frac{\partial^2 W}{\partial y^{*2}} f^2 + \frac{\partial W}{\partial y^*} f' f\right] + r^2 e^{-r(t^*-t)} \left[(1-\alpha) U_i - \alpha U_J\right] - 2 r e^{-r(t^*-t)} f \left[(1-\alpha) U_i' - \alpha U_J'\right] \\ & + e^{-r(t^*-t)} \left\{\left[\left(1 - \alpha U_i'' - \alpha U_J''\right)\right] f^2 + \left[(1-\alpha) U_i' - \alpha U_J'\right] f' f\right\} \end{aligned}$$

and

$$\begin{aligned} \frac{\partial^2 W}{\partial y^{*2}} f^2 + \frac{\partial W}{\partial y^*} f' f \quad = \quad & -r e^{-r(t^*-t)} g + e^{-r(t^*-t)} g' f + r^2 e^{-r(t^*-t)} \left[U_i + U_J\right] \\ & -2 r e^{-r(t^*-t)} \left[U_i' + U_J'\right] f + e^{-r(t^*-t)} \left\{\left[U_i'' + U_J''\right] f^2 + \left[U_i' + U_J'\right] f' f\right\} \end{aligned}$$

at $y^* = y_T^*$, where $f' = f'\left(y^*\right)$, $g' = g'\left(y^*\right)$. These two relations lead to

$$\alpha \left[-rg + g' f\right] + r^2 U_i + U_i' f \left[-2r + f'\right] + U_i'' f^2 \leq 0$$

at $y^* = y_T^*$. We can write the left-hand side as

$$-r \left[\alpha g - r U_i + U_i' f\right] + f \left[\alpha g - r U_i + U_i' f\right]'$$

Using (59) and assuming

$$f\left(y_T^*\right) > 0,$$

it follows that

$$\left[\alpha g\left(y^*\right) - r U_i\left(y^*\right) + U_i'\left(y^*\right) f\left(y^*\right)\right]' \leq 0 \tag{65}$$

at $y = y_T^*$. In terms of coefficients λ, σ, and γ, the left hand side is the same as

$$\frac{d}{dy^*}\left[\lambda\left(y^*\right)^2 + \sigma y^* + \gamma\right].$$

Hence,

$$2\lambda y_T^* + \sigma \leq 0.$$

This proves that the sign of the square root in (62) is negative.

The proof is complete.

Declarations

Ethical Approval This declaration is not applicable.

Competing interests All the co-authors declare no conflit of interest.

Authors' contributions The co-authors have contributed equally to the development of the research project and to the drafting of the resulting paper.

Funding Financial support from the French National Agency (grant ANR-17-EURE-0001) is gratefully acknowledged by Carmen Camacho.

Availability of data and materials This declaration is not applicable.

References

[1] Albrecht, J., A. Anderson, and S. Vroman (2010). Search by Committee. *Journal of Economic Theory*, 145(4), 1386-1407.

[2] Benchekroun, H., G. Gaudet, N. Van Long (2006). Temporary natural resource cartels. *Journal of Environmental Economics and Management*, 52, 663â€“674.

[3] 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.

[4] Bolton P., G. Roland and E. Spolaore (1996). Economic theories of the break-up and integration of nations. *European Economic Review*, 3(40), 697-705.

[5] 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.

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

[7] Boucekkine, R., C. Camacho, W. Ruan and B. Zou (2023). Optimal coalition splitting with heterogenous strategies. *Fulbright Review of Economics and Policy*. To appear.

[8] 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.

[9] 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.

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

[11] Colombo, L., P. Labrecciosa, N. Van Long (2022). A dynamic analysis of international environmental agreements under partial cooperation. *European Economic Review* 143, 104036.

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

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

[14] 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.

[15] Gancia, G., Ponzetto, G. A., and Ventura, J. (2020). A theory of economic unions. *Journal of Monetary Economics*, 109, 107-127.

[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 investment-related impacts of Brexit. *The World Economy*. 43, 2–32.

[18] Mayer, T., Vicard, V., and Zignago, S. (2019). The cost of non-Europe, revisited. *Economic Policy*, 34(98), 145-199.

[19] 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

[20] 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.

[21] Petrosjan, L. (1977). "Stable Solutions of Differential Games with Many Participants," Viestnik of Leningrad University 19, 46–52.

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

[23] Sampson T. (2017). Brexit: The Economics of International Disintegration. *Journal of Economic Perspectives* 31 (4), 163–184.

[24] Shiryaev, A.N. (2008). Optimal Stopping Rules. Springer-Verlag.

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

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

[27] 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.

[28] Van Long, N. (2010). A Survey on dynamic games in environmental economics. *Surveys on Theories in Economics and Business Administration*, 35-70. World Scientific.

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

[30] Zaccour, G. (2008). Time consistency in cooperative differential games: A tutorial. INFOR: Information Systems and Operational Research, 46(1), 81-92.

[31] 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, 1527–1567.

[32] Zhang, H. B., Dai, H. C., Lai, H. X., and Wang, W. T. (2017). US withdrawal from the Paris Agreement: Reasons, impacts, and China's response. *Advances in Climate Change Research*, 8(4), 220-225.

[33] Zou B. (2016). Differential games with (a)symmetric players and heterogeneous strategies. *Journal of Reviews on Global Economics* 5, 171-179.