CIRANO
Knowledge into action

A DYNAMIC ANALYSIS OF INTERNATIONAL ENVIRONMENTAL AGREEMENTS UNDER PARTIAL COOPERATION

LUCA **COLOMBO**
PAOLA **LABRECCIOSA**
NGO **VAN LONG**

2022s-01
WORKING PAPER

CS

CIRANO

Knowledge into action

Center for Interuniversity Research and Analysis on Organizations

ISSN 2292-0838 (online version)

A Dynamic Analysis of International Environmental Agreements under Partial Cooperation*

Luca Colombo[†] Paola Labrecciosa[‡] Ngo Van Long[§]

October 13, 2021

Abstract

We study the dynamics of equilibrium membership of an international environmental agreement aimed at increasing the stock of a global public good such as climate change mitigation. In contrast with previous studies, we assume partial cooperation among signatories, and show that the coalition size can be large, and increasing over time, even when the initial coalition size is small. We highlight a novel trade-off between agreements that are *narrow-but-deep-and-long-lived* vs. those that are *broad-but-shallow-and-short-lived*. We show that loose cooperative agreements, which are broad-but-shallow-and-short-lived, are both welfare- and Pareto-superior to tight cooperative agreements, which are narrow-but-deep-and-long-lived. We also show that conditions exist under which the equilibrium coalition size is efficient.

JEL Classification: C73; D60; H41; Q54.

Keywords: differential games; climate change mitigation; stable coalitions; coefficient of cooperation; social welfare.

*We thank Bård Harstad and Hassan Benchekroun for useful suggestions and comments. The usual disclaimer applies.

[†]Deakin Business School, Department of Economics, Burwood Campus, 221 Burwood Hwy, Burwood, 3125 VIC, Australia. Email: luca.colombo@deakin.edu.au.

[‡]Corresponding author. Monash Business School, Department of Economics, Clayton Campus, Wellington Road, Clayton, 3800 VIC, Australia. Email: paola.labrecciosa@monash.edu.

[§]McGill University, Department of Economics, 8551 Sherbrook St. W., Montreal, Quebec, H3A 2T7. Email: ngo.long@mcgill.ca.

Résumé

Nous étudions la dynamique d'adhésion à l'équilibre à un accord environnemental international visant à accroître le stock d'un bien public mondial tel que l'atténuation du changement climatique. Contrairement aux études précédentes, nous supposons une coopération partielle entre les signataires et montrons que la taille de la coalition peut être importante et augmenter au fil du temps même lorsque la taille initiale de la coalition est petite. Nous mettons en évidence un nouveau compromis entre les accords qui sont étroits mais profonds et de longue durée et ceux qui sont larges et superficiels mais de courte durée. Nous montrons que les accords de coopération partielle, qui sont larges mais superficiels et de courte durée, sont à la fois supérieurs en termes de bien-être aux accords de coopération serrés, qui sont étroits mais profonds et de longue durée. Nous montrons également qu'il existe des conditions dans lesquelles la taille de la coalition d'équilibre est efficace.

Mots-clés: jeux différentiels; atténuation du changement climatique; coalitions stables; coefficient de coopération; bien-être social.

Pour citer ce document / To quote this document

Colombo, N., Labrecciosa, P., Van Long, N. (2021). A dynamic analysis of international environmental agreements under partial cooperation (2022s-01).
https://www.cirano.qc.ca/fr/sommaires/2022s-01

1 Introduction

Many public goods are funded predominantly through voluntary contributions. This is especially the case for global public goods such as climate change mitigation, widespread peace, financial stability, and global public health. In the case of global public goods, the attainment of an efficient outcome requires international cooperation. However, experience from climate change policy indicates that full cooperation among all countries is hard to achieve. Partial cooperation seems to be a more realistic prospect due to conflicting national interests, disagreements on what constitutes a fair burden, and a general distrust among countries.

In the realm of climate change policy, it has recently been recognized that full cooperation could be less efficient than partial cooperation. A distinctly different approach from the "top-down" approach characterizing the Kyoto Protocol was taken at the Paris International COP21 Conference on Climate Change in 2015: countries agreed on an overall objective of limiting global warming to 2 degrees C relative to the pre-industrial temperature, but no country was required to set a specific target by a specific date. Indeed, unlike the Kyoto Protocol, the Paris Agreement does not insist on the coordination of climate actions among signatories: instead of setting commitments through centralized bargaining, the "bottom-up" approach of the Paris Agreement allows countries to make their own commitments.[1] There were some indications that this form of loose agreement, by attracting more participants, could turn out to be more effective in reducing emissions than the Kyoto Protocol.[2]

A well-known theoretical result in the literature on international environmental agreements is that the equilibrium coalition size is very small (except possibly when the potential benefits of

[1] Besides the Paris Agreement, many other IEAs can be classified as bottom-up agreements, which are generally based on unilateral pledges of mitigation action. A good example of a bottom up IEA is the Asia-Pacific Partnership on Clean Development and Climate (2006), adoped by Australia, Canada, China, India, Japan, Korea, and the US. One of the main objectives of this partnership is to create a voluntary, non-legally binding framework to facilitate the development and transfer of cleaner, more efficient technologies among partners. Another good example of a bottom up IEA is the Copenhagen Accord (2009). The pledges made in Copenhagen range from absolute targets with different base years over intensity and technology targets to sectoral mitigation activities.

[2] The reactions of the stock markets after the Paris Agreement was reached could be one such indication. Renewable energy share prices rose after the Paris Agreement. The iShare Global Clean Energy Exchange Trade Fund rose by 1.4% and the MAC Global Solar Energy index rose by 1.9%. Stock prices of coal companies fell sharply (11.3% for Peabody Energy, 4.9% for Consol Energy Inc.). The U.S. oil and gas index dropped by 0.5%. See Mukanjari and Sterner (2018), van der Ploeg and Rezai (2020).

cooperation are also small). This theoretical result is only partly supported by the fact that proposed agreements that prescribe full cooperation (e.g. the Kyoto Protocol) often fail to attract a sufficient number of truly committed signatories; it has been pointed out that the pessimistic conclusion reached in the literature that only agreements involving a very limited number of countries are stable is probably too pessimistic. Indeed, many real-world coalitions can be quite large.

Somewhat paradoxically, studies on international environmental agreements (or IEAs for short) that reach the overpessimistic conclusion about membership size were based on the overoptimistic assumption that signatories can be trusted to be fully committed to the maximization of joint welfare of members. In this paper, we propose and analyze a full-fledged continuous-time game of voluntary provision of a global public good with endogenous number of contributors that departs from the existing literature in a fundamental way. Instead of assuming full cooperation, we assume partial cooperation among signatory countries: each signatory country agrees to maximize a *weighted* sum of utilities of all members, giving greater weight to its own welfare. We believe that, with the exception of those situations characterized by negligible conflicts among nations, partial cooperation is a better description of reality than full cooperation. We model partial cooperation by using a coefficient of cooperation that can take any value from zero to one: a value of zero corresponds to the usual Nash behavior, and a value of one implies full cooperation (see Cyert and deGroot, 1973). If the designers of the IEA treaty specify that the coefficient of cooperation is in the interior of this range, it means that each signatory country is required to place a lower weight on the interest of other signatories than on its own interest. As pointed out in Harstad (2020b) with reference to the Paris Agreement: "Scholars and observers naturally expect a party's contribution to reflect that party's own interests to a larger extent, and the interests of other parties to a lower extent". Our modelling approach makes it possible to compare a Kyoto-style agreement, with a coefficient of cooperation close to one (meaning that countries are not allowed to place a lower weight on the interest of others) and a Paris-style agreement, with a coefficient of cooperation significantly lower than one (meaning that countries are allowed to place a lower weight on the interest of others).

The dynamic game at hand consists of a sequence of two-stage games. At each point in time there are two stages: the participation stage, followed by the contribution stage. At the participation stage, each country chooses independently and non-cooperatively whether or not to participate in the international agreement. For the equilibrium of the participation stage, we adopt the stability concept that is widely used in the literature on IEAs, requiring that a coalition be both internally and externally stable (see d'Aspremont et al., 1983).[3] At the contribution stage, each signatory country contributes to the public good the amount that maximizes its own utility plus a fraction (the coefficient of cooperation) of all the other members' utilities; each non-signatory country behaves as a Nash player by contributing to the public good the amount that maximizes its own utility. Countries are free to join or leave the agreement at any point in time.

Our equilibrium concept is Markov Perfect Equilibrium. We prove equilibrium existence and uniqueness. We show that if the exogenously-specified time horizon is finite (respectively, infinite) then the equilibrium size of the agreement increases (respectively, is constant) over time. In the case of a finite time horizon, we find that, in equilibrium, signatory countries continuously contribute to the public good until a certain (endogenously determined) point in time is reached, beyond which an agreement among contributing countries ceases to exist; non-signatory countries free-ride on the contributions made by the coalition. In this case, comparing with the social optimum, the equilibrium duration of the agreement is shorter than socially desirable, unless the coefficient of cooperation is unity. Moreover, we find that, since participation in an international agreement takes place on a voluntary basis, when the agreement is "too demanding", in the sense that it specifies joint utility maximization or a behavior close to it, participation is weak**.** When, instead, the agreement specifies a low coefficient of cooperation, participation is strong (i.e., the equilibrium coalition size tends to be larger). This is one of the key findings of our analysis. Intuitively, when the agreement is such that a country can place lower weight on the interest of others, it is not that costly for a country to participate, and this explains why the equilibrium coalition size is larger than in the case in which the agreement specifies joint-utility maximization. To our knowledge, the coefficient of cooperation has not been considered in the international agreement literature**,** neither in static nor in dynamic models. In the infinite-horizon case, we show that if the coefficient of cooperation is set at its lowest possible value, the equilibrium coalition size is efficient.

We add a third dimension, the time dimension, to the classical trade-off between agreements that are narrow-but-deep vs. those that are broad-but-shallow, thus enriching the comparative analysis

[3] Building on the work on cartel stability in oligopolistic markets by d'Aspremont et al. (1983), several studies have addressed the issue of the stability of IEAs, including Carraro and Siniscalco (1993), Barrett (1994), Rubio and Ulph (2007), and Karp and Simon (2013). An alternative set of stability criteria, referred to in the literature as "farsightedness", is considered in Diamantoudi (2005), de Zeeuw (2008), Benchekroun and Chaudhuri (2015), and Diamantoudi and Sartzetakis (2015).

of international agreements. We highlight a novel trade-off between agreements that are *narrow-but-deep-and-long-lived* vs. those that are *broad-but-shallow-and-short-lived*. We show that loose cooperative agreements are broad-but-shallow-and-short-lived, meaning that they entail strong participation but low cooperation and short (endogenous) duration, whereas tight cooperative agreements are narrow-but-deep-and-long-lived, meaning that they imply weak participation but high cooperation and long (endogenous) duration. We demonstrate that loose cooperative agreements are superior to tight cooperative agreements in terms of discounted welfare. To our knowledge, this is a novel result in the literature.[4] The policy implications of our results are clear: IEAs aimed at maximizing the welfare of its members as well as global welfare should be *designed* in a way to give more weight to individual rather than collective rationality. Relying exclusively on collective rationality usually leads to very small coalitions and inefficient outcomes.

1.1 Related Literature

A first strand of related literature studies dynamic games of IEAs involving *ex-ante* symmetric countries under the assumption that coalition members choose their policies to maximize their joint welfare, and non-members choose their policies to maximize their individual welfare (Battaglini and Harstad, 2016; Karp and Sakamoto, 2021; Kovác and Schmidt, 2021).[5] A common thread in this literature is that dynamic considerations can lead to large and long-lived agreements, thus solving the paradox typically arising from static analyses (e.g. Hoel, 1992; Carraro and Siniscalco, 1993; Barrett, 1994) that only small coalitions emerge when the potential gains from cooperation are large, a result known as the "paradox of international agreements" (Kolstad and Toman, 2005; Nordhaus, 2015).[6]

[4] On the possibility that IEAs be welfare-reducing under the assumption of full cooperation see Hoel (1992), Eichner and Pethig (2011), and Açıkgöz and Benchekroun (2017), *inter alia*.

[5] Battaglini and Harstad (2016), in particular, build on the early dynamic literature on IEAs, which includes Barrett (1994), Rubio and Ulph (2007), and de Zeeuw (2008).

[6] Karp and Simon (2013) shows that this paradox is sensitive to parametric assumptions. A solution to this paradox can also be found in the presence of trade sanctions (Hagen and Schneider, 2021), policy makers' reelection concerns (Battaglini and Harstad, 2020), trade measures such as border carbon adjustments (Khourdajie and Finus, 2020), parties that are troubled by time inconsistency (Gerlagh and Liski, 2018; Harstad, 2020c), or under ratification constraints (Köke and Lange, 2017). Other solutions to the above paradox are discussed in Rubio and Ulph (2006), allowing for Stackelberg leadership in the contribution stage, and in Finus and Maus (2008), considering a modesty parameter.

Battaglini and Harstad (2016) study a dynamic game of IEAs in which countries pollute and invest in green technologies. The length and depth of the agreement are endogenous: the coalition members negotiate the duration of the agreement and the abatement level for each participant. Two scenarios are considered, one in which the agreement specifies also the investments in green technologies (complete contracts), and one in which investments in green technologies are not contractible (incomplete contracts), leading to a hold-up problem that induces countries to invest little in green technologies unless the contract duration is sufficiently long. Battaglini and Harstad (2016) show that, in the incomplete contracting environment, large and long-lived agreements can arise in a Markov perfect equilibrium. Similarly to Battaglini and Harstad (2016), in our model, the length of the agreement is endogenous, being a function of the coalition size, and countries condition their strategies on state variables that evolve over time. Specifically, we consider a stock of public good increasing over time in response to countries' investments, whereas they consider a stock of pollution and a technology stock both increasing over time in response to countries' emissions and investments in green technologies, respectively. Unlike Battaglini and Harstad (2016), who assume concave benefits of consumption together with convex investment costs, we assume that both benefits of consumption and costs of contributing to the public good are linear.[7] Our modelling strategy allows us to derive clear-cut results on the dynamics of participation, which is a key new feature of our analysis. In particular, we are able to show that participation in an agreement increases over time (irrespective of the equilibrium size).[8]

Karp and Sakamoto (2021) analyze a dynamic stochastic game of IEAs in which a stable coalition is randomly selected by an endogenous probability distribution, which corresponds to players' beliefs. Karp and Sakamoto (2021) identify the connection between players' beliefs (e.g. their degree of optimism) and the stochastic process emerging from negotiation. Countries cannot commit to a coalition for more than a single period, and the (unique) state variable is the coalition inherited from the previous period. The members of today's coalition can decide at the beginning of the next period whether they maintain the coalition, or dissolve it. Unlike in our model, the equilibrium is a Markov chain (not a particular coalition). Karp and Sakamoto (2021) also establish an isomorphism

[7] Linear benefits and costs are also assumed in Barrett (1999), Barrett (2002), Ulph (2004), Kolstad (2011, Chapter 19), and Hong and Karp (2012).

[8] The opposite conclusion about the dynamics of participation is reached in Rubio and Ulph (2007). They study a dynamic model of IEAs in which the equilibrium coalition size decreases over time as the stock of a pollutant increases.

between their basic model without any stock of pollution, and an alternative version with a stock of pollution, and show that meaningful cooperation among countries requires an intermediate (i.e. sober) level of optimism. The results of their analysis are consistent with Battaglini and Harstad's (2016), in that small coalitions are abandoned in the next period, but the larger coalitions, once formed, are never abandoned (provided that countries are sufficiently patient). Similarly to Karp and Sakamoto (2021), we also show that dynamic considerations can lead to large and effective agreements. However, while Karp and Sakamoto (2021) adopt a stochastic membership approach, we adopt a deterministic one (as in Battaglini and Harstad, 2016). As such, in our model, countries beliefs about the success of an agreement play no role. Moreover, the focus of our analysis is different: while we focus on the dynamics of participation, Karp and Sakamoto (2021) focus on the conditions for an agreement to be sustainable, and abstract from public good dynamics.

Kovác and Schmidt (2021) analyze a dynamic stochastic game of IEAs that shares some features with that in Karp and Sakamoto (2021), in particular, the stochastic membership approach. Kovác and Schmidt (2021) modify the standard coalition formation game by assuming that, unlike in our model, countries can coordinate immediately on a long-term agreement, and that, if no long-term agreement is signed today, there is a delay and a new round of negotiations starts tomorrow. Such a delay is costly in the short-run, but may be profitable in the long-run if the countries anticipate that a better agreement can be signed in the future. The main insight from their analysis is that the sheer possibility of future negotiations can drastically change the outcome of the negotiations: in contrast with the static literature in which countries can negotiate only once, a large coalition that achieves substantial welfare gains can arise in equilibrium. Our paper differs from Kovác and Schmidt (2021) in two main respects, namely, our model is deterministic rather than stochastic, and the stock of public good continuously changes in response to countries' contributions over time (as long as an agreement exists), whereas Kovác and Schmidt (2021) abstract from public good dynamics.

Our paper contributes to the literature on dynamic games of IEAs discussed above by showing that dynamic considerations can lead not only to large and long-lived agreements, thus solving the "paradox of international agreements", but also to agreements that grow in size as time goes by, even when the initial size of the agreement is small. This is an important departure from the existing literature, in which long-lived agreements are sustainable if and only if the equilibrium size of the agreement is large. Indeed, one of the key results of our analysis is that, in the finite-horizon case, in line with the stylized facts about the dynamics of many real-world IEAs (see Section 5), participation increases over time irrespective of the initial size of the agreement.

Most of the literature on IEAs (e.g. Hoel, 1992; Carraro and Siniscalco, 1993; Barrett, 1994; Rubio and Ulph, 2007; de Zeeuw, 2008; Eichner and Pethig, 2013; Karp and Simon, 2013; Battaglini and Harstad, 2016; Karp and Sakamoto, 2021; Kovác and Schmidt, 2021) assumes that signatory countries agree to "give up their own self interest" and act in a way to maximize the sum of all members' utilities. A variation to the standard setup of full cooperation among signatory countries is considered in Van der Pol et al. (2012), Hoel and de Zeeuw (2014), and Buchholtz et al. (2014).

Van der Pol et al. (2012) extend the standard two-stage game for the analysis of IEAs by including altruism in the participation decision. Two types of altruism are analyzed: impartial altruism, where countries show a concern for all other countries, and community altruism, where the concern extends only to coalition partners. Altruism is captured by means of two non-negative parameters, one that multiplies the sum of utilities of all members of the coalition, and one that multiplies the sum of utilities of all outsiders: community (resp. impartial) altruism exists when the value of the former parameter exceeds (resp. equals) the value of the latter. A key finding of their analysis is that, compared with a scenario in which countries are motivated only by their own wellbeing, a scenario in which countries have altruistic preferences is associated with larger coalitions. Van der Pol et al. (2012) also show that even a small degree of altruism is sufficient to stabilize the Grand Coalition. Our approach in modelling partial cooperation is closely related to theirs (albeit the interpretation of coefficients is quite different).[9] However, Van der Pol et al. (2012) do not explore any dynamic issues.

Hoel and de Zeeuw (2014) consider a three-stage game of IEAs with a participation stage (Stage 1), an R&D stage (Stage 2), and an emission stage (Stage 3). In Stage 1, each country decides whether or not to join an agreement in which participating countries undertake R&D activities aimed to reduce abatement costs. In Stage 2, the coalition decides how much to invest in R&D (and how to share this cost among its members). Finally, in Stage 3, all countries (coalition members and outsiders) decide how much to abate. Hoel and de Zeeuw (2014) assume that countries may cooperate through research joint ventures on the development of new, climate friendly technology that reduces the costs of abatement (Stage 2 cooperation), and that there is no coop-

[9] In modelling partial cooperation, our approach shares also some features with that taken by Harstad (2020a), where each contributor to a public good maximizes an *asymmetric* Nash product where the weight on others' payoffs is smaller than in the standard Nash Bargaining Solution.

eration on emission reductions (Stage 3 competition). They show that such partially cooperative agreements can lead to lower emissions compared with those resulting from fully cooperative agreements. Partial cooperation, as modelled in Hoel and de Zeeuw (2014), has no direct counterpart in our framework. Nevertheless, one can think of our sequence of two-stage games as a sequence of reduced-form games where a coefficient of cooperation lower than one captures the fact that Stage 1 payoffs do not correspond to those derived under the assumption of full cooperation in every stage of the game.

Buchholtz et al. (2014), by adopting an aggregative game approach, analyze a two-stage game to investigate the implications of partial cooperation for the provision of global public goods. Two different groups of countries coexist: one is a coalition of like-minded cooperating countries whose members are mutually matching their public good provision, and the other consists of outsiders which, without any matching, act non-cooperatively playing Nash against the coalition. Matching rates are set in the first stage of the game, then national contributions are chosen in the second stage. Buchholtz et al. (2014) show that, counterintuitively, an increase in the number of countries may lead to a reduction of equilibrium public good supply. In their model, partial cooperation among coalition participants is modelled through reciprocal matching of public good contributions within the coalition, which is an important difference with respect to our way of modelling partial cooperation. In addition, in their model, dynamic considerations are absent.

Our paper also contributes to the literature on the well-known trade-off between the depth and the breadth of cooperation, i.e., between agreements that are *narrow-but-deep* vs. those that are *broad-but-shallow* (see Schmalensee, 1998; Barrett, 2002; Aldy et al., 2003; Finus and Maus, 2008; Harstad, 2020b), by showing that there exists a negative relationship between the depth of cooperation, measured, in our model, by a coefficient of cooperation ranging from zero (for no cooperation) to one (full cooperation), and the breadth of cooperation, measured by the number of signatory countries. As previously argued, we add another dimension to the above trade-off, giving rise to a new trade-off between agreements that are *narrow-but-deep-and-long-lived* vs. those that are *broad-but-shallow-and-short-lived*.

Finally, our paper relates to the literature on voluntary provision of public goods in dynamic settings, which includes Fershtman and Nitzan (1991), Wirl (1996), Marx and Matthews (2000), Itaya and Shimomura (2001), Yanase (2006), Benchekroun and Long (2008), Fujiwara and Matsueda

(2009), Battaglini et al. (2014), Georgiadis (2015, 2017), and Bowen et al. (2019), *inter alia*.[10]

The remainder of this paper is organized as follows. The game theoretical model is laid down in Section 2. Section 3 derives the social optimum. Section 4 characterizes the equilibrium of the game. Section 5 provides some evidence on the dynamics of IEAs. Section 6 contains a welfare analysis. Section 7 discusses the assumptions and possible extensions. Section 8 concludes.

2 The Game

The game is specified in continuous time. Time is denoted by $t \in [0, T\rangle$. The time horizon, T, is exogenously specified. It can be finite or infinite. There are $n \geq 2$ *a priori* identical countries. The game at hand consists of a sequence of two-stage games. At each $t \in [0, T\rangle$, there are two stages: (i) a participation stage; (ii) a contribution stage. In stage (i), each country, motivated only by self-interest, decides independently and non-cooperatively whether or not to participate in an IEA that aims at increasing the stock of a global public good such as climate change mitigation. In stage (ii), the contribution of each participating country is decided by the coalition under the terms of the treaty, whereas non-participating countries decide how much to contribute independently and non-cooperatively. This is as in the canonical IEA model (e.g. Barrett 1999; Hong and Karp, 2012). The main departure from the existing dynamic literature is that the coalition requires each participant to contribute an amount that maximizes the discounted value of a weighted sum of utilities of all participants rather than the discounted value of the sum of utilities of all participants. In line with the existing dynamic literature, non-participating countries act as Nash players, each with the aim of maximizing its own discounted utility. The agreement exists until time $\widehat{T} \leq T$, with $\widehat{T}$ being *endogenously* determined. A country's decision to join the agreement at any $t_0 \in [0, \widehat{T}\rangle$ has to be rational at that time (in terms of discounted sum of utilities), but each country realizes that it may become irrational for it to remain in the agreement at some $t_1 \in [0, \widehat{T}\rangle$, with $t_1 > t_0$. If that turns out to be the case, the country will leave the agreement at t_1. Similarly, a country's decision to stay out of the agreement at any $t_0 \in [0, \widehat{T}\rangle$ has to be rational at that time; joining the agreement may become rational at $t_1 \in [0, \widehat{T}\rangle$, with $t_1 > t_0$. For the sake of simplicity, we assume that countries are free to join or leave the agreement at any $t \in [0, \widehat{T}\rangle$ without having to pay any entry/exit cost. Let

[10] Classical references on voluntary provision of public goods in static settings include Chamberlin (1974), Morrison (1978), Bergstrom at al. (1986), Bernheim and Douglas (1986), Cornes and Sandler (1986), Andreoni (1988), and Varian (1994).

$m(t) \leq n$ denote the number of signatories at t ($n - m(t)$ denotes the number of non-signatories); signatories are denoted by the index $i = 1, ..., m$, and non-signatories by the index j. For each country, the utility is linear in the stock of public good, $K(t)$.[11] Without any loss in generality, we normalize the benefit of each unit of additional K to 1. The public good is entirely financed with countries' contributions, with $x_k(t)$ denoting country k's contribution at t, $k = i, j$. Each country's total cost of contributing is $cx_k(t)$, with $c > 0$. Further restrictions on c will be specified later on (see Assumption A2). The restrictions imposed on $x_k(t)$ are given in Assumption A1.

Assumption A1. $x_k(t) \in [0, 1]$ for all $t \in [0, T\rangle$.

Assumption A1 implies that each country's contribution cannot be negative and cannot exceed its maximum capacity, normalized to 1 throughout the game.

When country i participates in an IEA, we distinguish country i's material payoff from the objective function that is used by the coalition to determine country i's contribution levels (as long as country i remains a signatory). Country i instantaneous material payoff is

$$u_i(t) = K(t) - cx_i(t), \tag{1}$$

and its contribution levels result from the maximization of the present value of the stream of weighted sum of instantaneous material payoffs of all members of the agreements, $v_i(t)$, defined as

$$v_i(t) = u_i(t) + \phi_i \sum_{k \neq i}^{m(t)} u_k(t), \tag{2}$$

where $u_k(t)$ is defined analogously to (1) and $\phi_i \in [0, 1]$ denotes the coefficient of cooperation (see Cyert and deGroot, 1973), specified in the agreement, with $i, k = 1, ..., m(t)$, $k \neq i$ (i.e. country k is a signatory country other than i).[12] Note that $\phi_i = 1$ means that country i is required to fully internalize the benefit that its contribution x_i confers on all other signatories. Therefore $\phi_i = 1$ *corresponds to the case of full internalization (joint utility maximization)*, whereas any $\phi_i \in (0, 1)$ describes partial internalization. When $\phi_i = 0$, countries act non-cooperatively. We assume that $\phi_i = \phi$ for all $i = 1, ..., m(t)$. Those countries entering an agreement are required to contribute to

[11] This assumption is common in the literature on voluntary provision of public goods (or, equivalently, abatement of public bads). See, for instance, Barrett (1999), Hong and Karp (2012), and Battaglini et al. (2014).

[12] The coefficient of cooperation was called the coefficient of "effective sympathy" by Edgeworth (1881). For recent applications of the coefficient of cooperation to environmental economics and industrial organization see Colombo and Labrecciosa (2018) and Lopez and Vives (2019), respectively.

the public good so as to maximize the present value of the stream of (2). For agreements specifying a high ϕ we use the expression tight cooperative agreements; for agreements specifying a low ϕ we use the expression loose cooperative agreements.

The representative non-signatory country j cares only about its own material payoff:

$$u_j(t) = K(t) - cx_j(t). \tag{3}$$

The main difference between (2) and (3) is that while signatories agree to cooperate (for any $\phi \in (0,1]$), non-signatories choose to act in isolation.

The evolution of the stock of public good is governed by the following differential equation:

$$\frac{dK(t)}{dt} = X(t), \; K(0) = K_0 \geq 0,$$

where $X(t)$ denotes the sum of all contributions at t.

Let $r > 0$ be the discount rate, the same for all countries. In determining country i's contribution level, the coalition solves the following problem:

$$\begin{cases} \max\limits_{x_i(t)} J_i = \int\limits_0^T e^{-rt} v_i(t)\, dt \\ s.t. \; x_i(t) \in [0,1] \text{ and } \frac{dK(t)}{dt} = x_i(t) + X_{-i}(t), \end{cases} \tag{4}$$

where X_{-i} is the sum of all contributions except the contribution made by country i, and v_i (defined in (2)) corresponds to the weighted sum of instantaneous material payoffs of all members of the agreements. The contribution of each signatory country is determined so as to maximize the discounted value of the weighted sum of instantaneous material payoffs of all members of the agreements subject to the feasibility constraint on the contribution levels and the evolution of the stock of public good. When choosing $x_i(t)$, the coalition values the investment in the durable stock K not only in terms of the future benefit stream that country i will reap, but also a fraction ϕ of the future benefit stream that other signatories will reap.

Country j's contribution solves the following problem:

$$\begin{cases} \max\limits_{x_j(t)} J_j = \int\limits_0^T e^{-rt} u_j(t)\, dt \\ s.t. \; x_j(t) \in [0,1] \text{ and } \frac{dK(t)}{dt} = x_j(t) + X_{-j}(t), \end{cases} \tag{5}$$

where X_{-j} is the sum of all contributions except the contribution made by country j, and u_j (defined in (3)) corresponds to the instantaneous material payoff of a non-signatory country. Each non-signatory country chooses a level of contribution to the public good so as to maximize the discounted value of its own material payoff subject to the feasibility constraint on the contribution levels and the evolution of the stock of public good. When country j chooses $x_j(t)$, it only values such an investment in terms of its own future benefit stream.

We are interested in characterizing a Markov Perfect Equilibrium of the dynamic game at hand. We assume that countries' strategies are of the Markovian type: their optimal decision rules on contribution levels at t are functions of the current stock K, the current date, t, and the current participation status of all players, as denoted by $\mathbf{p}(t) = \{p_1(t), ..., p_n(t)\}$, where $p_z \in \{in, out\}$, with $z = 1, ..., n$. Thus, contribution strategies can be written as $x_k(t) = \sigma_k(K, t, \mathbf{p}(t))$, where σ_k denotes the Markovian decision rule, with $k = i, j$. In equilibrium, contribution strategies are denoted by $x_k^*(t)$. The participation status at t is determined by a Markovian strategy $p_z(t) = \psi_z(K, t)$, where ψ_z denotes the decision rule, with $z = 1, ..., n$.

We stress two important assumptions. First, in line with the bulk of the existing literature on IEAs, for non-participating countries, contributions are strategic, but for each individual participating country, contributions are non-strategic, in that they are determined by the coalition. Second, all countries are selfish: the coefficient of cooperation (specified in the agreement) has nothing to do with altruism. It only affects the level of contributions of signatories. When countries decide to join or leave the agreement, they do so purely based on self-interest: they use only their material payoff function (1) to evaluate the desirability to join or to leave.

In order to sharpen our results, we make the following assumption on the cost parameter c.

Assumption A2. *The cost parameter is strictly greater than $\underline{c}$ and strictly smaller than $\bar{c}$, where $\underline{c} = 1/r$ and $\bar{c} = n(1 - e^{-rT})/r$, and T is sufficiently long such that $n(1 - e^{-rT}) > 1$.*

Our specification that $c > \underline{c}$ implies that if a country acts in isolation (i.e., not participating in the agreement), it will find that its individually rational contribution level is zero. The second specification, that $c < \bar{c}$, implies that, from the social welfare perspective, the social benefit (in terms of sum of utilities for all countries) of adding a unit of capital outweighs the cost of investment (at least when the end of the time horizon is far away).

For future reference, we define instantaneous global welfare at t as

$$w(t) = nK(t) - c\left\{m(t)\, x_i^*(t) + [n - m(t)]\, x_j^*(t)\right\},$$

which can be used to compute discounted global welfare,

$$W = \int_0^T e^{-rt} w(t)\, dt. \tag{6}$$

Instantaneous global welfare is equal to the difference between total benefits and total costs at t. Total benefits are given by nK, given that n corresponds to the total number of countries (sum of signatories and non-signatories) and that the benefit of each unit of additional K is normalized to 1. Total costs are given by sum of the cost of contribution borne by signatories, cmx_i^*, and the cost of contribution borne by non-signatories, $c(n-m)x_j^*$.

3 The Social Optimum Benchmark

In this section, we derive the socially optimal level of contributions to the public good. The socially optimal solution will be used as a benchmark against which one can compare the equilibrium outcomes of the game under various specifications of the coefficient of cooperation, ϕ.

The social planner's problem can be written as

$$\begin{cases} \max\limits_X W = \int_0^T e^{-rt} \left[nK(t) - cX(t) \right] dt \\ s.t.\ X \in [0, n] \text{ and } \frac{dK(t)}{dt} = X(t), \end{cases} \tag{7}$$

where X denotes the aggregate level of contribution to the public good. The social planner chooses how much to contribute to the public good so as to maximize the discounted sum of instantaneous welfare subject to the feasibility constraint on the contribution levels and the evolution of the stock of public good. Note that the feasibility constraint that the social planner faces is different from the one in country k's maximization problem, with $k = i, j$, the reason being that in a world populated by n countries, each of which can contribute at most 1, the upper bound of contributions equals n.

Proposition 1 *The socially optimal level of contributions to the public good is given by*

$$X^{so}(t) = \begin{cases} n & \text{for } t \in [0, \widetilde{T}\rangle \\ 0 & \text{otherwise,} \end{cases}$$

with

$$\widetilde{T} = \begin{cases} T + \frac{1}{r} \log\left(1 - \frac{cr}{n}\right) \leq T & \text{if } T \text{ is finite} \\ \infty & \text{if } T \text{ is infinite.} \end{cases}$$

Proof. See Appendix A. ■

Proposition 1 establishes that, if the time horizon T is finite, it is socially desirable to stop contributing after time $\widetilde{T}$: when the terminal date approaches, the shadow price of the public good becomes smaller than the cost of contributing. If, instead, the time horizon is infinite, then contributions are constant and equal to $X^{so}(t) = n$ at any point in time, because the discounted value of an infinite stream of benefits of an additional unit of capital is n/r which is greater than the cost c, as $c < \overline{c} = n(1 - e^{-rT})/r = n/r$ for $T = \infty$.

4 The Equilibrium of the Game

We proceed as follows: first, we determine the contribution levels at t for a given number of participants; then, we endogenize the number of participants at t; finally, we check whether the equilibrium conditions for the two-stage game at t are satisfied for all two-stage games, i.e., for all $t \in [0, T\rangle$, or they are satisfied only for a time interval within $[0, T\rangle$.

4.1 The Contribution Stage at t

By standard arguments, Markov Perfect Equilibrium strategies must satisfy the following HJB equations. For a signatory country,

$$\begin{aligned} rV_i(K,t) &= \max_{x_i \in [0,1]} \left\{ K - cx_i + \phi \sum_{k \neq i, k=1}^{m(t)} [K - c\sigma_k(K,t,\mathbf{p}(t))] + \frac{\partial V_i(K,t)}{\partial K} \right. \\ &\quad \left. \times \left[x_i + \sum_{k \neq i, k=1}^{m(t)} \sigma_k(K,t,\mathbf{p}(t)) + \sum_{j=m(t)+1}^{n} \sigma_j(K,t,\mathbf{p}(t)) \right] + \frac{\partial V_i(K,t)}{\partial t} \right\}, \qquad (8) \end{aligned}$$

and, for a non-signatory country,

$$\begin{aligned} rV_j(K,t) &= \max_{x_j \in [0,1]} \left\{ K - cx_j + \frac{\partial V_j(K,t)}{\partial K} [x_j + \sum_{i=1}^{m(t)} \sigma_i(K,t,\mathbf{p}(t)) \right. \\ &\quad \left. + \sum_{k \neq j, k=m(t)+1}^{n-1} \sigma_k(K,t,\mathbf{p}(t))] + \frac{\partial V_j(K,t)}{\partial t} \right\}, \qquad (9) \end{aligned}$$

with $i = 1, ..., m(t)$, $j = m(t)+1, ..., n$. $\partial V_i(K,t)/\partial K$ in (8) and $\partial V_i(K,t)/\partial K$ in (9) denote the shadow price of the public good in the maximization of problem of country i and j, respectively. The shadow price represents the amount a country is willing to pay for a marginal increase in the stock of public good. Maximization of (8) implies that[13]

$$x_i^* = \begin{cases} 1 \text{ if } \frac{\partial V_i(K,t)}{\partial K} \geq c \\ 0 \text{ otherwise,} \end{cases} \tag{10}$$

whereas maximization of (9) implies that

$$x_j^* = \begin{cases} 1 \text{ if } \frac{\partial V_j(K,t)}{\partial K} \geq c \\ 0 \text{ otherwise.} \end{cases} \tag{11}$$

Not surprisingly, country $k = i, j$ will contribute as long as the shadow price of the public good exceeds or is equal to the marginal cost of contribution.[14] The shadow prices in (10) and (11) are decreasing over time, reaching the value of zero at the terminal date (see Appendix B). This implies that for t sufficiently close to T there will be no contributions to the public good.

We will use the above information to determine whether a signatory (respectively, non-signatory) country would be better off to deviate, changing its status to non-signatory (respectively, signatory). The deviation decision is based solely on a country's self-interest, i.e., only a country's present and future material payoffs (u_i and u_j) matter, and the function v_i plays no role in this calculation. The (Stage 2) equilibrium material payoff of a generic country $k = i, j$ is given by the discounted value of instantaneous payoffs, with instantaneous payoffs given by (2) for $k = i$ (i.e. for signatories) and by (3) for $k = j$ (i.e. for non-signatories). The expressions of equilibrium payoffs are given in Appendix B.

4.2 The Participation Stage at t

In equilibrium, two groups of players can be identified: a group $G_1(t)$ consisting of $m(t)$ signatories (where $m(t)$ is endogenously determined), and a group $G_2(t)$ consisting of $n - m(t)$ non-signatories. For a (Stage 1) equilibrium with $m(t)$ signatories to exist it must be that each $i \in G_1(t)$ has no unilateral incentive to deviate and join $G_2(t)$, and each $j \in G_2(t)$ has no unilateral incentive to deviate by joining $G_1(t)$. Formally, for each $i \in G_1(t)$, it must hold that

$$\Pi_i (K, t; m(t)) \geq \Pi_j (K, t; m(t) - 1), \tag{12}$$

and for each $j \in G_2(t)$, it must hold that

$$\Pi_j (K, t; m(t)) \geq \Pi_i (K, t; m(t) + 1), \tag{13}$$

[13] We assume that if $\frac{\partial V_i(K,t)}{\partial K} = c$ then signatories will use the tie-breaking rule that $x_i^* = \sup [0, 1]$.

[14] We assume that if $\frac{\partial V_j(K,t)}{\partial K} = c$ then non-signatories will use the tie-breaking rule that $x_j^* = \sup [0, 1]$.

for all $K \geq 0$ and $t \in [0, T\rangle$. Inequality (12) states that, for a participating country, it must be individually rational to participate in the agreement (in which case the total number of participating countries is m) rather than acting in isolation.[15] We will refer to condition (12) as the **contributor-rationality** condition. (12) corresponds to the internal stability condition in the coalition literature.

Inequality (13) states that, for a non-signatory, it must be better to stay out of the agreement rather than joining group $G_1(t)$ (in which case it assumes that the number of signatories becomes $m+1$, and that all of them contribute at their new symmetric Nash equilibrium level with $m+1$ signatories). We will refer to condition (13) as the **free-rider-rationality** condition. (13) corresponds to the external stability condition in the coalition literature. In line with the bulk of the literature, we assume that (13) needs to hold with strict inequality sign, i.e. a country which is indifferent between joining and not joining will join.

4.3 Equilibrium Characterization

We are now in a position to characterize the equilibrium of the game. It turns out that, given the parameter value of ϕ (the coefficient of cooperation), there exists a corresponding threshold level $\widehat{c}$ such that, if the cost parameter c is below $\widehat{c}$, we can determine both (i) the endogenous time $\widehat{T} < T$ at which all countries cease to contribute to the public good (if T is finite); and (ii) the equilibrium number of signatories at t, $m^*(t)$, for each $t \in [0, \widehat{T}\rangle$. For ease of reference, let us state the threshold level $\widehat{c}$ below. Define

$$\widehat{c} = \frac{[(n-1)\,\phi + 1]\left(1 - e^{-rT}\right)}{r}. \tag{14}$$

Note that (i) $\widehat{c} \leq \overline{c}$, with equality holding iff $\phi = 1$, and that (ii) $\widehat{c} > \underline{c}$ if T is sufficiently large.

Proposition 2 *Let $\widehat{c} \leq \overline{c}$ be defined by equation (14). Assume T is sufficiently large so that $\widehat{c} > \underline{c}$. Then for all $c \in (\underline{c}, \widehat{c}]$, and for all $t \in [0, \widehat{T}\rangle$, with*

$$\widehat{T} = \begin{cases} T + \frac{1}{r}\log\left(1 - \frac{cr}{1+\phi(n-1)}\right) < T & \text{if } T \text{ is finite} \\ \infty & \text{if } T \text{ is infinite} \end{cases}$$

[15] Each country assumes that if it leaves group G_1 to join group G_2, then (i) the number of signatories will become $m-1$, i.e., the other $m-1$ signatories stay in G_1; and (ii) the remaining signatories will adjust their contribution level to the Nash equilibrium level with $m-1$ signatories.

there exists a unique equilibrium agreement, and the number of signatory countries at time t, denoted by $m^*(t)$, is given by

$$2 \leq m^*(t) = f\left(1 + \frac{1}{\phi}\left(\frac{cr}{1 - e^{r(t-T)}} - 1\right)\right) \leq n, \tag{15}$$

where $f(.)$ is an integer-valued function that maps any real number y to the smallest integer that is greater than or equal to y. The equilibrium level of contributions to the public good is given by

$$X^*(t) = \begin{cases} m^*(t) & \text{for } t \in [0, \widehat{T}\rangle \\ 0 & \text{otherwise.} \end{cases}$$

Proof. See Appendix B. ■

Proposition 2 derives the threshold of c below which there exists an agreement among m^* countries contributing to the public good. $\widehat{c}$ given in (14) corresponds to the highest possible value of the shadow price of the public good in each signatory country's maximization problem. When c is larger than $\widehat{c}$, it becomes too costly for each signatory country to contribute, therefore an agreement is never reached. $\widehat{c}$ is increasing in ϕ, n, T, and decreasing in r, implying that the likelihood of cooperative agreements to exist is higher the tighter the agreement (i.e., the higher ϕ), the higher the number of countries which could potentially participate in the agreement, the longer the time horizon of each country, or the less each country discounts future payoffs. When an agreement among contributing countries does not arise in equilibrium, the stock of public good remains at its initial level, K_0, for all $t \in [0, T\rangle$. This is so because the shadow price of the public good is higher for signatories than for non-signatories, and $c > \widehat{c}$ implies that it is also individually irrational for each non-signatory country to contribute. Note that, in the finite-horizon case, $\widehat{T} < \widetilde{T}$, i.e. in equilibrium, countries will stop contributing to the public good earlier than under social planning, the reason being that the shadow price of the public good is higher for the social planner, who fully internalizes the impact that a country's contribution has on other countries, than for any coalition among partially cooperating countries (including the grand coalition). In the infinite-horizon case, $\widehat{T} = \widetilde{T} = \infty$. In this case, the private contributions to the public good are always positive, albeit lower than the socially optimal ones. Interestingly, for $c \in (\widehat{c}, \overline{c})$, the private contributions to the public good are nil, despite the fact that it would be socially desirable to contribute to the public good. Intuitively, a dynamic free-riding problem arises: each signatory country values the investment in the durable stock K in terms of the future benefit stream that it will reap plus a fraction ϕ of the future benefit stream that other signatories will reap, whereas the social planner values the the investment in the durable stock K in terms of the future benefit stream that all countries (signatories and non-signatories) will reap. Interestingly, for $t \in [0, \widehat{T}\rangle$, $X^*(t)$ increases over time, therefore free-riding diminishes over time. However, for $t \in (\widehat{T}, \widetilde{T}\rangle$, $X^*(t) = 0$ while $X^{so}(t) = n$, thus free-riding is at its extreme. We can then argue that, provided that $c < \widehat{c}$, free riding is non-monotone with respect to time. If, instead, $c \geq \widehat{c}$ then free riding will be constant (and maximal) over time: $X^*(t) = 0$ while $X^{so}(t) = n$ for all $t \in [0, \widetilde{T}\rangle$.

Remark 1 *Partial cooperation is responsible for free-riding to be non-monotone with respect to time.*

When $\phi = 1$, we have $\widehat{T} = \widetilde{T}$, therefore the duration of the agreement in equilibrium turns out to be socially optimal. Despite this, compared with social optimum, there is under-contribution. The intuitive explanation is that while the social planner makes it compulsory for countries to participate in the agreement, in the unregulated scenario, countries' participation is voluntary, and there are private incentives for countries to free ride on the contributions to the public good made by signatories, thus leading to a lower participation than that which is socially optimal. Similar to the case in which $\phi < 1$, for $t \in [0, \widehat{T}\rangle$, when $\phi = 1$, free riding decreases over time until reaching zero at $\widehat{T}$. In the final phase, for $t \in (\widehat{T}, T\rangle$ investment by the social planner is zero, and thus we can say that free-riding is nil. Hence, if $\phi = 1$**,** free-riding will be always non-increasing over time.

Corollary 1 *An agreement among $2 \leq m^*(t) \leq n$ contributing countries specifying a coefficient of cooperation $\phi < \underline{\phi}$ does not exist, with*

$$\underline{\phi} = \frac{e^{-rT} + cr - 1}{(n-1)\left(1 - e^{-rT}\right)}.$$

If $T \geq \underline{T}$ then $\underline{\phi} \in (0, 1]$, with

$$\underline{T} = \frac{1}{r} \log\left(\frac{n}{n - cr}\right).$$

If $T < \underline{T}$ then $\underline{\phi} > 1$ and an agreement among $2 \leq m^*(t) \leq n$ contributing countries does not exist.

Proof. See Appendix C. ■

The above corollary implies that if the time horizon is very short ($T < \underline{T}$) then voluntary public good provision is not an equilibrium outcome of the game, even if signatories agree to fully cooperate. In the social optimum, instead, even when the time horizon is very short, it is always optimal for the social planner to provide the public good (given Assumption A2).

Corollary 2 *In the finite-horizon case, a loose cooperative agreement exists for a shorter time than a tight cooperative agreement.*

Proof. See Appendix D. ■

Corollary 2 can be restated as follows: the time it takes to reach universal participation is shorter for loose than for tight cooperative agreements. This is so because, interpreting $m^*(t)$ as a continuous variable, the rate of change in participation is given by (with over-dot denoting time derivative)

$$\frac{\dot{m}^*(t)}{m^*(t)} = \frac{cr^2 e^{rt}}{\left(e^{rT} - e^{rt}\right)\left[cr - (1-\phi)\left(1 - e^{r(t-T)}\right)\right]}, \tag{16}$$

which, for any finite T, is positive (by Assumption A2) and decreasing in ϕ. Hence, it takes less time to reach universal participation for agreements specifying a low value of ϕ (i.e. loose cooperative agreements) than for agreements specifying a high value of ϕ (i.e. tight cooperative agreements). Note that the rate of change in participation is positive for all values of ϕ, including 1, in which case (16) becomes $re^{rt}/(e^{rT} - e^{rt})$.

From (16), we can state the following corollary.

Corollary 3 *In the finite-horizon case, the equilibrium size of the agreement increases over time, irrespective of the initial size of the agreement.*

When instead T is infinite, m^* becomes independent of time and equal to $f(1 + (cr - 1)/\phi)$.

The result that participation in IEAs increases over time is in line with the dynamics of participation in several real-world IEAs. Indeed, our model is capable of accounting for an important well-documented fact about international environmental cooperation (see Section 5). The intuition why, in our model, participation is increasing over time is as follows. The equilibrium size of the agreement is such that, at any point in time, the shadow price of the public good is equal to the marginal cost of contributing to the public good. While the former is constant, the latter is decreasing over time, since, in the finite-horizon case, as time goes by, there is less and less time for countries to reap the benefits from the public good. The shadow price of the public good is increasing in the number of participating countries, since there exists cooperation, at least partial, among them, implying that each signatory country takes into account the benefit stream that other signatories will reap. For the equilibrium condition to hold at ant $t \in [0, \widehat{T}]$ the size of the agreement has to increase over time (until reaching universal participation).

Corollary 4 *The equilibrium size of the agreement is larger for a loose cooperative agreement than for a tight cooperative agreement.*

Proof. The proof is immediate since m^* is decreasing in ϕ (for $d\phi$ sufficiently large). ■

Intuitively, an increase in the coefficient of cooperation increases the likelihood of a positive contribution (from (10)), thus leading to an increase in the incentive to free ride. The result that participation in an international agreement is larger for loose cooperative agreements than to tight cooperative agreements can explain the larger participation in a loose cooperative agreement such as the Paris Agreement than in a tight cooperative agreement such as the Kyoto Protocol.

Loose cooperative agreements are more successful in attracting participation by being less ambitious. The logic resembles that behind partial collusion in repeated games: when the critical discount factor sustaining full collusion is too high compared with the discount factor used by firms to discount future profits, it is still possible for firms to sustain some degree of cooperation rather than behaving as Nash players provided that firms' discount factor is sufficiently high.

In the light of Corollaries 2 and 4, we are in a position to establish a novel trade-off between **narrow-but-deep-and-long-lived** agreements vs. **broad-but-shallow-and-short-lived** ones. (The expressions "narrow-but-deep" and "broad-but-shallow" are characterizations used in the literature on IEAs; see Harstad 2020b and references therein.)

Proposition 3 *A loose cooperative agreement is broad-but-shallow-and-short-lived (i.e. it implies strong participation but low cooperation and short duration) whereas a tight cooperative agreement is narrow-but-deep-and-long-lived (i.e. it implies weak participation but high cooperation and long duration).*

Proposition 3 adds a third dimension, the time dimension, to the classical trade-off between agreements that are narrow-but-deep vs. broad-but-shallow. In the static counterpart of our model, there exists a unique trade-off, that between participation in the agreement and cooperation in the agreement. In our dynamic model, instead, there are three trade-offs: (i) a trade-off between participation in the agreement and cooperation in the agreement (the standard static trade-off); (ii) a trade-off between participation in the agreement and duration of the agreement; (iii) a trade-off between cooperation in the agreement and duration of the agreement. Compared with the static literature on narrow-but-deep vs. broad-but-shallow agreements, we provide a richer comparative analysis of international agreements. Our framework yields results that are coherent with well-established facts about international cooperation that cannot be accounted for in static models, such as increasing participation over time, and shed new light on the social desirability of international cooperation. Which of the two cooperative agreements, loose or tight, should be favored on welfare grounds? Is efficiency achievable, and, if so, under what circumstances? We will tackle these questions in Section 6.

We conclude this section with the impact of c and r on the equilibrium membership size at time t, $m^*(t)$. Using equation (15), the following comparative statics properties hold: (i) $m^*(t)$ is non-decreasing in c, either constant, if dc is such that $f(y)$ remains unchanged, or increasing. Intuitively, an increase in c increases the incentive to "share the burden", thus leading to bigger participation. (ii) $m^*(t)$ is non-decreasing in r, either constant, if dr is such that $f(y)$ remains unchanged, or increasing. Intuitively, an increase in the discount rate decreases the likelihood of a positive contribution (from (10)), thus leading to a decrease in the incentive to free ride. Note that $m^*(t)$ is independent of n if $m^*(t) < n$, otherwise it is increasing in n.

5 The Dynamics of IEAs: Some Evidence

In this section, we report some stylized facts about real world IEAs, with particular emphasis on the evolution of participation in the agreements. Data have been collected from the International Environmental Agreements Database Project (available at https://iea.uoregon.edu/) for six IEAs, namely, the Paris Agreement (2015), the Kyoto Protocol (1997), the Copenhagen Amendment (1992), the Montreal Protocol (1987), the Convention on International Trade in Endangered Species of Wild Flora and Fauna (CITIES, 1973), and the World Heritage Convention (1972). The evolution of participation in these IEAs is depicted in Figure 1.

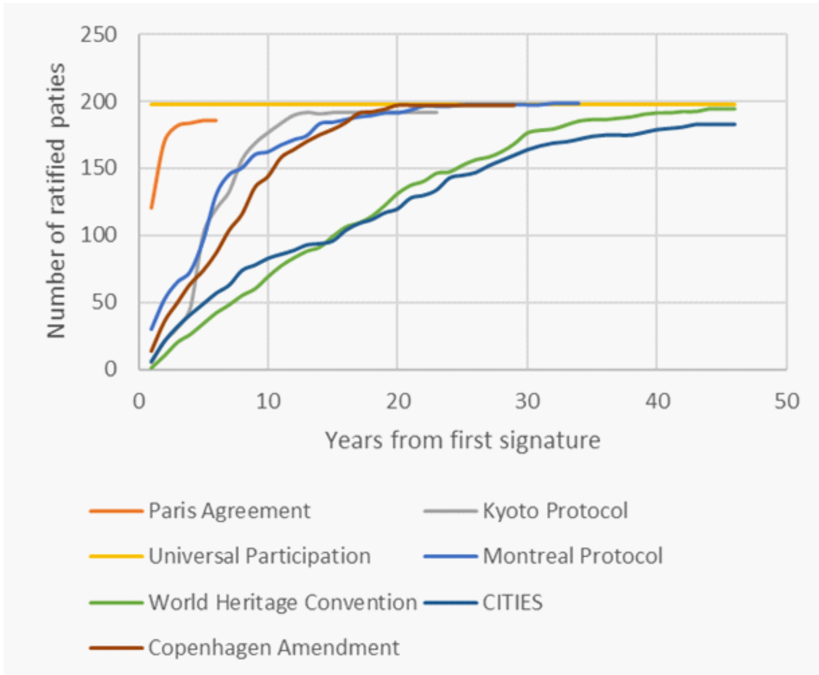

Figure 1: The dynamics of IEAs

From Figure 1, four stylized facts can be inferred.

1. Participation in IEAs is increasing over time (irrespective of initial size). For instance, the number of ratified parties increased from 1 to 194 in the World Heritage Convention, and from 5 to 192 in the Kyoto Protocol.

2. After some years from first signature, IEAs are large in size. For instance, the number of ratified parties is equal to 183 in CITIES (after 43 years from first signature), and to 197 in the Montreal Protocol (after 25 years from first signature).

3. At any point in time, participation is higher the higher the initial number of ratified parties. For instance, after 10 years from first signature, the number of ratified parties is equal to 69 in the World Heritage Convention, and to 177 in the Kyoto Protocol.

4. The rate of change in participation is higher the higher the initial number of ratified parties, which implies that the time it takes to reach (almost) universal participation is shorter the higher the initial number of ratified parties. For instance, after five years from first signature,

the number of ratified parties increased from 5 to 101 in the Kyoto Protocol, and only from 1 to 34 in the World Heritage Convention.

All the above four stylized facts are broadly consistent with our theory, which predicts that participation is increasing over time (see Corollary 3 for Fact 1), reaching a relatively large size after a short time period (see Proposition 2 for Fact 2), and that both participation and the rate of change in participation are increasing in the initial number of ratified parties, at any point in time (see Corollary 4 for Fact 3 and Corollary 2 for Fact 4).

From Proposition 2, we know that the initial number of ratified parties is decreasing in the coefficient of cooperation, ϕ. As argued in Section 1, ϕ can be assumed to be higher for more centralized, top-down agreements such as the Kyoto Protocol than for more decentralized, bottom-up agreements such as the Paris Agreement. In line with Figure 1, our theory predicts that for both agreements, participation is relatively large after a short time period, and that the initial number of ratified parties, participation, and the rate of change in participation at any point in time are higher for the Paris Agreement than for the Kyoto Protocol.

As can be seen in Figure 1, the dynamics of the Montreal Protocol is similar to that of the Kyoto protocol and also to that of the Copenhagen Amendment, and the dynamics of the World Heritage Convention is similar to that of CITIES. We can then identify three groups of IEAs, Group 1, composed by the Paris Agreement in isolation, Group 2, composed by the Montreal Protocol, the Kyoto Protocol, and the Copenhagen Amendment, and Group 3, composed by the World Heritage Convention and CITIES. We can assign a different value of ϕ to each group: the lowest value to Group 1, the highest value to Group 3, and an intermediate value to Group 2. Assigning the highest value to Group 3 is justifiable given that, by their nature, IEAs in Group 3 require full cooperation.[16] The findings from the comparison among Groups 1,2, and 3, are consistent with those from the comparison between the Paris Agreement and the Kyoto Protocol. Over time, a shift in approach from the centralized, top-down agreements signed in the 70s (e.g. CITIES) to the

[16] The World Heritage Convention was adopted in 1972 and came into force in 1975. It aims to promote cooperation among nations to protect cultural and natural heritage around the world that is of such outstanding universal value that its conservation is important for current and future generations. States that are parties to the Convention agree to identify, protect, conserve, and present World Heritage properties, and do all they can with their own resources to protect their World Heritage properties. CITIES was adopted in 1973 and came into force in 1975. It aims to ensure that international trade in listed species of wild animals and plants does not threaten their survival in the wild. CITES is legally binding and requires parties to implement national legislation to enforce its requirements.

decentralized, bottom-up agreements signed in more recent years (e.g. the Paris Agreement) can be observed. Participation and the rate of change in participation (after any given number of years from first signature) are clearly lower for earlier agreements (which, in our theory, are cooperative agreements specifying a high value of ϕ) than for more recent ones (which, in our theory, are cooperative agreements specifying a low value of ϕ).[17]

6 Welfare Analysis of IEAs

In this section, we compare and contrast loose and tight cooperative agreements in terms of discounted global welfare, and show that conditions exist under which efficiency can be achieved.

Proposition 4 *At any point in time $t \in [0, \widehat{T}\rangle$, discounted global welfare is higher under a loose than under a tight cooperative agreement.*

Proof. See Appendix E. ■

The intuitive explanation behind Proposition 4 is as follows. From Corollary 4, we know that an increase in ϕ leads to a decrease in participation in the agreement, m^*. In Appendix E, we show that an increase in ϕ also leads to a decrease in the stock of public good at any point in time, K^*. *Ceteris paribus*, while the latter is clearly a welfare-reducing effect, the former is a welfare-enhancing effect (as long as the agreement exists), since a decrease in participation in the agreement leads to a decrease in the total cost of contribution to the public good. Proposition 4 establishes that the cost-saving effect from decreased participation is always outweighed by the negative effect that an increase in ϕ has on the stock of public good, thus making discounted global welfare during an IEA higher under a loose than under a tight cooperative agreement.

Proposition 4 does not consider the welfare after the end of an agreement (which we provide in Appendix E). The impact of ϕ on such welfare turns out to be ambiguous, depending on the parameter values. A numerical analysis reveals that for c small and T large the welfare-superiority of loose cooperative agreements extends also to $t \in (\widehat{T}, T\rangle$. Note that the difference between T and $\widehat{T}$ shrinks as n increases. A larger n implies that the period after the end of an agreement shortens,

[17] At COP26 to be held in Glasgow in Novemeber 2021 countries must finalize the Paris Rulebook (the rules needed to implement the Paris Agreement). In this respect, Glasgow 2021 can be considered as a continuation of the bottom-up approach undertaken in the Paris Agreement.

and for $n \to \infty$ we have $\widehat{T} \to T$. Hence, we expect the result in Proposition 4 to hold for $t \in [0, T\rangle$ provided that n be sufficiently large.

As $T \to \infty$, we have $\widehat{T} \to \infty$ (from Proposition 2), and W converges to (when evaluated at $t = 0$)

$$W = \frac{rnK_0 + (n - cr)\, m^*}{r^2}.$$

Interestingly, the result in Proposition 4 does not depend on the initial stock of public good, K_0. This is so because K_0 does not affect participation in the agreement.

Proposition 4 is illustrated by means of the following numerical example.

Numerical Example. Let $r = 0.1$, $K_0 = 0$, $c = 12$, $n = 120$, $T = 10$. We consider two scenarios: (i) $\phi = 0.8$ (tight cooperative agreement - Kyoto style); (ii) $\phi = 0.2$ (loose cooperative agreement - Paris style). It turns out that $\widehat{T} = 9.87448$ in (i) and $\widehat{T} = 9.50403$ in (ii). By Proposition 4, we expect $W|_{\phi=0.2} > W|_{\phi=0.8}$ at any $t \in [0, 9.50403]$, i.e. given that an IEA exists in both scenarios. Discounted global welfare evaluated at $\phi = 0.2$ and at $\phi = 0.8$ are depicted in Figure 2 below.

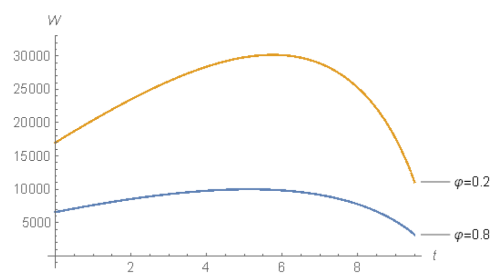

Fig. 2: Welfare comparison

Since both Π_i and Π_j are increasing in m, loose cooperative agreements are not only welfare-superior but also Pareto-superior.

Corollary 5 *Let $T \to \infty$. If $\phi = \underline{\phi}$ then $m^* = n$ for all $t \in [0, \infty)$. In this case, the equilibrium size of the agreement and the contributions to the public good are both socially optimal.*

When $T \to \infty$ (therefore the difference between $\widehat{T}$ and $\widetilde{T}$ converges to zero) discounted global welfare is maximized by setting $\phi = \underline{\phi}$, with $\underline{\phi}$ denoting the lower bound of ϕ for an agreement to exist (see Corollary 1). Any increase in the coefficient of cooperation above $\underline{\phi}$ leads to a reduction in discounted global welfare by reducing the equilibrium number of contributors (unless $d\phi$ is such that m^* remains unchanged) and, in turn, the stock of public good. Interestingly, efficiency is achieved despite the fact that signatory countries do not fully coordinate. $\underline{\phi}$ is increasing in c and r and decreasing in n. Not surprisingly, discounted global welfare is increasing in n. As to the impact of c, in the infinite-horizon case, we have (when evaluated at $t = 0$)

$$\frac{\partial W}{\partial c} = \frac{(n - cr)}{r^2}\frac{\partial m^*}{\partial c} - \frac{m^*}{r}.$$

Since $\partial m^*/\partial c \geq 0$ and $c < n/r$ (by Assumption A2) *a priori*, $\partial W/\partial c$ can be either positive or negative. Counterintuitively, a small increase in c can be welfare-improving because it can lead to a discrete increase in m^* which outweighs the negative direct effect. When instead $dc > 0$ is sufficiently large, despite a discrete increase in m^*, discounted global welfare decreases.

When T is finite, the difference between $\widehat{T}$ and $\widetilde{T}$ converges to zero as $\phi \to 1$. The duration of the agreement is indeed socially optimal in the case of joint utility maximization. However, in this case, there exists underprovision of the public good w.r.t. the social optimum, since $m^* < n$ for all $t \in [0, \widehat{T})$. Therefore, when T is finite, efficiency cannot be achieved. As to the impact of c, in the finite-horizon case, the logic is the same as that in the infinite-horizon case: a small increase in c can be welfare-improving.

7 Discussion of Assumptions and Possible Extensions

Given the complexity of the problems we study, it is not surprising that our model is based on several simplifying assumptions, which could be relaxed in future works. First of all, we abstract from asymmetry among countries. There are at least four possible types of asymmetry that could be considered. Asymmetry in: (i) the cost of contributing to the public good; (ii) the benefit from the public good; (iii) capacity, i.e. in the maximum amount that a country can contribute to the public good; (iv) the coefficient of cooperation among countries. As to (i), an alternative assumption to $c_i = c$ for all countries, could be that there are at least two groups of countries differentiated on the basis of the marginal cost of contributing to the public good, e.g. Groups 1 and 2, with $c_1 > c_2$, which could be the case of developed (Group 2) and developing countries (Group 1). As to (ii), instead of normalizing the marginal benefit of the public good to 1 for all countries, the analysis could be extended to the case in which this marginal benefit differs across countries. As to (iii), instead of assuming that $x_i \in [0,1]$ for all countries, one could assume that $x_i \in [0,1]$ with country i belonging to the group of small countries, and $x_j \in [0,\lambda]$ for country j belonging to the group of big countries, with $\lambda \geq 1$ capturing the asymmetry in the size of the economy among countries. As to (iv), instead of assuming $\phi_i = \phi$ for all countries, one could think of two (or more) groups of countries differentiated on the basis of their willingness to cooperate, e.g. Groups 1 and 2, with $\phi_1 > \phi_2$, which could be the case of developed and developing countries with the former being more willing (and more able) to take the welfare of others into account than the latter. By continuity, we expect our results to hold for small asymmetries.

Another simplifying assumption made in our analysis concerns the exogeneity of the coefficient of cooperation, ϕ. In a (richer and more realistic) model in which ϕ is endogenously determined, either once and for all at the beginning of the game or at each stage of the game, it is reasonable to assume that ϕ would stem from some sort of bargaining process in which larger countries have higher bargaining power and are therefore in a position to leverage their circumstances to strike more desirable deals with other countries. For instance, one could endogenize the coefficient of cooperation by maximizing the product of surplus utilities (Nash, 1950), or by equalizing the ratios of maximal gains (Kalai-Smorodinsky, 1975), or by maximizing the minimum of surplus utilities (Kalai, 1977). The difficulty with endogenizing the coefficient of cooperation in a dynamic setting such as ours is that the choice made by the coalition at $t = 0$ should be time-consistent, i.e. it should remain optimal throughout the entire planning horizon: even if it were possible for signatories to renegotiate the terms of the agreement over time countries should have no unilateral incentives to deviate from the initially set level of ϕ. Related to this, it would be interesting to extend the analysis to allow for the terms of the agreement to change over time by assuming that ϕ itself is a function of time.

A further simplifying assumption made in our analysis is that both benefits and costs are linear. In an attempt to keep the model as simple as possible, the same assumption has been made in other papers studying coalition formation (e.g. Barrett, 1999; Hong and Karp, 2012). We build on the canonical (static) model of IEAs by introducing public good dynamics, with the aim to obtain clear-cut analytical results on the dynamics of participation (which is the main focus of our analysis). Linearity in benefits and costs implies corner solutions, rendering the decision about public good contribution a binary one. A nonlinear model such as that in Battaglini and Harstad (2016) would make the analysis of public good contribution more involved, probably jeopardizing clear-cut analytical results about the dynamics of participation.

Finally, in our analysis, we do not examine investments in green technologies and uncertainty, and we do not take depreciation into account. Modifying the equation of motion for the public good to incorporate a random term such as a white noise and (constant) capital depreciation would not significantly change the results of our analysis. As to the impact of uncertainty, neither equilibrium strategies nor payoffs (in expected terms) would be affected, given the linearity of the value functions w.r.t. the stock of public good. As to capital depreciation, say δ, as long as δ is small, participation in the agreement and the stock of public good would remain monotonically increasing over time until the end of the agreement (in the finite-horizon case). For large δ we would expect a steady-state solution to arise, in which case participation in the agreement and the stock of public good would reach a plateau. Including a second state variable and investments in green technologies (as done in Battaglini and Harstad, 2016) would enrich the model, making it suitable to extend the analysis to the holdup problem in the presence of coalition formation.

8 Concluding Remarks

We have proposed and analyzed a multi-country continuous-time game of voluntary provision of a global public good such as climate change mitigation. Our dynamic game consists of a sequence of two-stage games. At each point in time, there are two stages: (i) a participation stage; (ii) a contribution stage. In the first stage, each country decides independently and non-cooperatively whether or not to join the agreement. This decision is based solely on self-interest. In the second stage, the contribution level of each participating country is determined by the coalition with the aim of maximizing a weighted sum of utilities of all participants, whereas each non-participating country decides how much to contribute to the public good, independently and non-cooperatively, with the aim of maximizing its own utility. The assumption of partial rather than full cooperation among participating countries represents one of the main departures from the existing literature on coalition formation and IEAs. This seemingly small departure has important consequences for equilibrium membership and welfare.

Our analysis has shown that a loose cooperative agreement, specifying a low coefficient of cooperation, is associated with larger participation than a tight cooperative agreement, specifying a high degree of cooperation. In contrast with the conventional wisdom according to which the equilibrium coalition size is small and typically inefficient, we have shown that loose cooperative agreements can lead to an equilibrium coalition size that is large and efficient. We have shown that discounted global welfare is higher under loose than under tight cooperative agreements; in the infinite-horizon case, efficiency is achieved when the coefficient of cooperation is set at its lower bound. A policy implication of this finding is that insisting on coordination among voluntary contributors to a public good is generally welfare reducing. In the realm of climate change policy, a loose agreement in the style of the Paris Accord is likely to be more successful than its predecessor, the Kyoto Protocol, which proved to be "too demanding" for countries to join.

A particularly important contribution of our paper is the addition of a third dimension, the time dimension, to the classical trade-off between agreements that are narrow-but-deep vs. broad-but-shallow. In our dynamic game, strong participation has to be weighted against not only low cooperation but also short duration of the agreement. This leads to a novel trade off between agreements that are narrow-but-deep-and-long-lived vs. broad-but-shallow-and-short-lived.

In contrast with previous studies on dynamic voluntary provision of public goods (e.g. Battaglini and Harstad, 2016; Karp and Sakamoto, 2021; Kovác and Schmidt, 2021), we have shown that relatively small coalitions can become bigger over time, which is in line with what has been observed in relation to several real-world IEAs.

Appendix A. Proof of Proposition 1

We solve the problem by making use of the Hamilton-Jacobi-Bellman (HJB) equation

$$rV(K,t) = \max_{X\in[0,n]}\left\{nK - cX + \frac{\partial V(K,t)}{\partial K}X + \frac{\partial V(K,t)}{\partial t}\right\}, \tag{A.1}$$

where $V(K,t)$ is the value function to be determined. Maximization of the right-hand side of (A.1) implies that[18]

$$X^*(t) = \begin{cases} n \text{ if } \frac{\partial V(K,t)}{\partial K} \geq c \\ 0 \text{ otherwise.} \end{cases}$$

To determine the value function, we guess a value function of the form $V(K,t) = A(t)K + B(t)$, where $A(t)$ and $B(t)$ are to be solved for. Consider first the interval of time such that $A(t) \geq c$, in which $X^* = n$. From (A.1), it follows that

$$r\left[A(t)K + B(t)\right] = n\left[K - c + A(t)\right] + \frac{\partial A(t)}{\partial t}K + \frac{\partial B(t)}{\partial t},$$

which implies that, over this phase, we must have

$$A(t) = \frac{n\left[1 - e^{r(t-T)}\right]}{r}, \tag{A.2}$$

and

$$B(t) = \frac{n\left\{e^{r(t-T)}\left\{cr + n\left[r(t-T) - 1\right]\right\} - cr + n\right\}}{r^2},$$

where we have used the boundary condition $A(T) = B(T) = 0$. Since $A(t)$ is decreasing in t, then $A(t) \geq c$ for $t \in [0, \widetilde{T}\rangle$, with $\widetilde{T}$ solving $A(t) = c$. Note that $A(t)$ attains its highest value at $t = 0$, with $A(0) = n(1 - e^{-rT})/r$. By Assumption A2, we have $A(0) > c$. Then, near the end of the time horizon, i.e., for $t > \widetilde{T}$, the optimal investment is zero, $X^* = 0$. From (A.1), it follows that, when $X^* = 0$,

$$r\left[A(t)K + B(t)\right] = nK + \frac{\partial A(t)}{\partial t}K + \frac{\partial B(t)}{\partial t}, \text{ for all } t \in [\widetilde{T}, T],$$

which implies that over the interval $[\widetilde{T}, T]$, the function $A(t)$ is also given by the equation (A.2) which holds for $t < \widetilde{T}$, while $B(t) = 0$ for $t \in [\widetilde{T}, T]$.

When T is finite, we have $X^* = n$ as long as $t \in [0, \widetilde{T}]$ and $X^* = 0$ for $t \in (\widetilde{T}, T]$. When instead T is infinite, we have $X^* = n$ for all $t \in [0, \infty)$, given that $\lim_{T\to\infty} \widetilde{T} = \infty$.

[18] We assume that if $\frac{\partial V(K,t)}{\partial K} = c$ then the social planner will use the tie-breaking rule that $X^* = \sup[0, n]$.

Appendix B. Proof of Proposition 2

The proof proceeds in three steps. In Step 1, we derive the (Stage 2) value functions and equilibrium payoffs. In Step 2, we prove that $m^*(t)$ is an equilibrium. Then, in Step 3, we show that it is unique.

Step 1. Assume $V_k(K,t) = A_k(t)K + B_k(t)$, with $k = i, j$, and $A_k(T) = B_k(T) = 0$. There are four sub-cases to consider.

1a. $\partial V_i(K,t)/\partial K \geq c$ and $\partial V_j(K,t)/\partial K \geq c$. In this sub-case, $x_i^* = x_j^* = 1$. Value functions for the signatories are given by

$$V_i = A_i(t)K + B_i(t),$$

where

$$A_i(t) = \frac{[\phi(m-1)+1]\left(1-e^{r(t-T)}\right)}{r},$$

and

$$B_i(t) = \frac{[\phi(m-1)+1]\left\{e^{r(t-T)}\{cr + n[r(t-T)-1]\} + (n-cr)\right\}}{r^2},$$

and value functions for the non-signatories are given by

$$V_j = A_j(t)K + B_j(t),$$

where

$$A_j(t) = \frac{1-e^{r(t-T)}}{r},$$

and

$$B_j(t) = \frac{e^{r(t-T)}\{cr + n[r(t-T)-1]\} + (n-cr)}{r^2}.$$

Since $A_j(t) < \underline{c} = 1/r$ then this sub-case does not exist, being in conflict with the assumption that $c > \underline{c}$.

1b. $\partial V_i(K,t)/\partial K \geq c$ and $\partial V_j(K,t)/\partial K < c$. In this sub-case, $x_i^* = 1$ and $x_j^* = 0$. Value functions for the signatories are given by

$$V_i = A_i(t)K + B_i(t),$$

where $A_i(t)$ is as in sub-case 1a, and

$$B_i(t) = \frac{[\phi(m-1)+1]\left\{e^{r(t-T)}\{cr + m[r(t-T)-1]\} + (m-cr)\right\}}{r^2},$$

and value functions for the non-signatories are given by

$$V_j = A_j(t) K + B_j(t),$$

where $A_j(t)$ is as in sub-case 1a, and

$$B_j(t) = \frac{m \left\{ e^{r(t-T)} [r(t-T) - 1] + 1 \right\}}{r^2}.$$

This sub-case exists when $\underline{c} < c$ and $[\phi(m-1)+1]\left(1 - e^{r(t-T)}\right)/r \geq c$. The first inequality is satisfied by Assumption A2; the second inequality holds true if $m \geq m^*(t)$, with $m^*(t)$ being specified by eq. (15).

1c. $\partial V_i(K,t)/\partial K < c$ and $\partial V_j(K,t)/\partial K \geq c$. In this sub-case, $x_i^* = 0$ and $x_j^* = 1$. Value functions for the signatories are given by

$$V_i = A_i(t) K + B_i(t),$$

where $A_i(t)$ is as in sub-case 1a, and

$$B_i(t) = \frac{(n-m)[\phi(m-1)+1] \left\{ e^{r(t-T)} [r(t-T) - 1] + 1 \right\}}{r^2},$$

and value functions for the non-signatories are given by

$$V_j = A_j(t) K + B_j(t),$$

where $A_j(t)$ is as in sub-case 1a, and

$$B_j(t) = \frac{e^{r(t-T)} \{cr + (n-m)[r(t-T) - 1]\} + n - m - cr}{r^2}.$$

This sub-case cannot exist. (It would exist only if both $[\phi(m-1)+1]\left(1 - e^{r(t-T)}\right)/r < c$ and $[1 - e^{r(t-T)}]/r > c$ hold, which is impossible.)

1d. $\partial V_i(K,t)/\partial K < c$ and $\partial V_j(K,t)/\partial K < c$. In this sub-case, $x_i^* = x_j^* = 0$. Value functions for the signatories are given by

$$V_i = A_i(t) K + B_i(t),$$

where $A_i(t)$ is as in sub-case 1a, and

$$B_i(t) = 0,$$

and value functions for the non-signatories are given by

$$V_j = A_j(t) K + B_j(t),$$

where $A_j(t)$ is as in sub-case 1a, and

$$B_j(t) = 0.$$

This sub-case exists when $[\phi(m-1)+1]\left(1-e^{r(t-T)}\right)/r < c$, which holds if $m < m^*(t)$, with $m^*(t)$ being specified by eq. (15).

To sum-up, only sub-cases 1b and 1d are possible. The (Stage 2) equilibrium payoffs turn out to be

$$\Pi_i(K,t;m^*) = \begin{cases} \frac{m^*+r(K-c)-e^{r(t-T)}\{r(K-c)+m^*[r(T-t)+1]\}}{r^2} \text{ for } t \in [0,\widehat{T}\rangle \\ \frac{\left(1-e^{r(t-T)}\right)K}{r} \text{ for } t \in (\widehat{T},T\rangle \end{cases} \tag{B.1}$$

and

$$\Pi_j(K,t;m^*) = \begin{cases} \frac{m^*+rK-e^{r(t-T)}\{rK+m^*[r(T-t)+1]\}}{r^2} \text{ for } t \in [0,\widehat{T}\rangle \\ \frac{\left(1-e^{r(t-T)}\right)K}{r} \text{ for } t \in (\widehat{T},T\rangle. \end{cases}$$

with $\widehat{T}$ being the unique value of t that solves $\left. A_i(t)\right|_{m=n} = c$.

Step 2. Consider first the case where

$$f\left(1+\frac{1}{\phi}\left(\frac{cr}{1-e^{r(t-T)}}-1\right)\right) < n,$$

so that there are $m^* < n$ members. From (12) and (13), we have that m^* is an equilibrium if

$$\begin{aligned} \Pi_i(K,t;m^*) &= \frac{m^*+r(K-c)-e^{r(t-T)}\{r(K-c)+m^*[r(T-t)+1]\}}{r^2} \\ &\geq \Pi_j(K,t;m^*-1) = \frac{\left(1-e^{r(t-T)}\right)K}{r}, \end{aligned} \tag{B.2}$$

and

$$\begin{aligned} \Pi_j(K,t;m^*) &= \frac{m^*+rK-e^{r(t-T)}\{rK+m^*[r(T-t)+1]\}}{r^2} \\ &> \Pi_i(K,t;m^*+1) = \frac{m^*+1+r(K-c)-e^{r(t-T)}}{r^2} \\ &\quad \times \{r(K-c)+(m^*+1)[r(T-t)+1]\}. \end{aligned} \tag{B.3}$$

Inequality (B.2) can be rewritten as

$$\Pi_j(K,t;m^*) - \Pi_j(K,t;m^*-1) = \frac{e^{r(t-T)}\{cr-m^*[r(T-t)+1]\}+m^*-cr}{r^2} \geq 0,$$

which is decreasing in t for $m^* > c/(T-t)$ and nil at $t=T$. $m^* - c/(T-t)$ is increasing in c, and nil at $c = c_1$, with

$$c_1 = \frac{(1-\phi)(T-t)\left(e^{r(t-T)}-1\right)}{\phi\left(1-e^{r(t-T)}\right)-r(T-t)} < \frac{1}{r}.$$

Recall that $c > 1/r$ by Assumption A2. Therefore, $m^* > c/(T-t)$. This proves that $\Pi_i(K,t;m^*) > \Pi_j(K,t;m^*-1)$.

Inequality (B.3) can be rewritten as

$$\Pi_j(K,t;m^*) - \Pi_i(K,t;m^*+1) = \frac{e^{r(t-T)}[1 - r(c+t-T)] + cr - 1}{r^2},$$

which is increasing in c and nil at $c = c_2$, with

$$c_2 = (T-t)\left(1 + \frac{1}{e^{r(t-T)} - 1}\right) + \frac{1}{r} < \frac{1}{r}.$$

Therefore, $\Pi_j(K,t;m^*) > \Pi_i(K,t;m^*+1)$.

Consider next the case where

$$f\left(1 + \frac{1}{\phi}\left(\frac{cr}{1 - e^{r(t-T)}} - 1\right)\right) = n,$$

so that there are $m^* = n$ members. From (12) we have that $m^* = n$ is an equilibrium if

$$\begin{aligned}\Pi_i(K,t;n) &= \frac{n + r(K-c) - e^{r(t-T)}\{r(K-c) + n[r(T-t)+1]\}}{r^2} \\ &\geq \Pi_j(K,t;n-1) = \frac{\left(1 - e^{r(t-T)}\right)K}{r}.\end{aligned}$$

By the same logic as that used to prove that $m^* > c/(T-t)$ we have that $\Pi_i(K,t;n) > \Pi_j(K,t;n-1)$. Note that, since $m^* = n$, the free-rider-rationality condition does not apply. Note also that when $m^* = n$ the equilibrium level of contributions is nil.

Step 3. Consider m^*+1 members, with $m^* \leq n-1$. We have that m^*+1 is an equilibrium if

$$\begin{aligned}\Pi_i(K,t;m^*+1) &= \frac{m^*+1 + r(K-c) - e^{r(t-T)}\{r(K-c) + (m^*+1)[r(T-t)+1]\}}{r^2} \\ &\geq \Pi_j(K,t;m^*) = \frac{m^* + rK - e^{r(t-T)}\{rK + m^*[r(T-t)+1]\}}{r^2},\end{aligned}$$

which can be rewritten as

$$\Pi_i(K,t;m^*+1) - \Pi_j(K,t;m^*) = \frac{e^{r(t-T)}[r(c+t-T) - 1] - cr + 1}{r^2} \geq 0.$$

Since $c > 1/r$ by Assumption A2 then $\Pi_i(K,t;m^*+1) < \Pi_j(K,t;m^*)$ implying that m^*+1 does not satisfy the contributor-rationality condition: the extra member would find it rational to leave. Analogously, consider m^*+k, with $k \geq 1$ and $m^* \leq n-k$. We have

$$\Pi_i(K,t;m^*+k) < \Pi_j(m^*+k-1).$$

Hence, any outcome with more than $m^* < n$ members does not satisfy the contributor-rationality condition. Consider now $m^* - 1$ members, with $m^* < n$. We have

$$\begin{aligned} \Pi_j\left(K,t;m^*-1\right) &= \frac{\left(1-e^{r(t-T)}\right)K}{r} < \Pi_i\left(K,t;m^*\right) = \frac{m^*+r\left(K-c\right)-e^{r(t-T)}}{r^2} \\ &\times\left\{r\left(K-c\right)+m^*\left[r\left(T-t\right)+1\right]\right\}. \end{aligned}$$

Therefore, $m^* - 1$ does not satisfy the free-rider-rationality condition: the first excluded member would find it rational to join. Finally, consider $m^* - k$, with $k \geq 1$ and $m^* < n$. We have

$$\Pi_i\left(K,t;m^*-k\right) = \Pi_j^*\left(K,t;m^*-k-1\right).$$

Under the tie-breaking assumption that a country which is indifferent between joining and not joining will join, any outcome with less than $m^* < n$ members does not satisfy the free-rider-rationality condition. We can then conclude that there exists a unique m^*, with m^* given by equation (15).

It is immediate to verify that $\widehat{T}$ is the value of t that solves

$$1+\frac{1}{\phi}\left(\frac{cr}{1-e^{r(t-T)}}-1\right)=n,$$

and that $\widehat{T} \in (0,T)$ for $c < \widehat{c}$. For $c \geq \widehat{c}$ we have either $\widehat{T} < 0$ or $\widehat{T} \notin \mathbb{R}$. As a consequence, since

$$\frac{\left[\phi\left(n-1\right)+1\right]\left(1-e^{r(t-T)}\right)}{r} < c,$$

then $m^*(t) = n$ and $X^*(t) = 0$ for all $t \in [0, T\rangle$.

Appendix C. Proof of Corollary 1

The inequality $c \leq \widehat{c}$ in Proposition 2 can be solved for ϕ to get $\underline{\phi}$, which reduces to $(cr-1)/(n-1)$ as T tends to infinity. If $\phi > \underline{\phi}$ then $m^*(t) < n$; if $\phi = \underline{\phi}$ then $m^*(t) = n$; if $\phi < \underline{\phi}$ then $c > \partial V_i(K,t)/\partial K$ implying that $X^*(t) = 0$. $\underline{\phi}$ is decreasing in T and $\underline{\phi} = 1$ at $\underline{T}$. Therefore, $\underline{\phi} \leq (>) 1$ for $T \geq (<) \underline{T}$.

Appendix D. Proof of Corollary 2

We have

$$\frac{\partial\widehat{T}}{\partial\phi}=\frac{c\left(1-n\right)}{\left[\left(n-1\right)\phi+1\right]\left[cr-1-\left(n-1\right)\phi\right]}>0,$$

since $c < \widehat{c}$, with $\widehat{c}$ given in Proposition 2.

Appendix E. Proof of Proposition 4

From (6) and Proposition 2, discounted global welfare turns out to be

$$W = \begin{cases} \frac{\left(1-e^{r(t-T)}\right)nK^*}{r} + \frac{\left\{e^{r(t-T)}\{cr+n[r(t-T)-1]\}+n-cr\right\}m^*}{r^2} & \text{for } t \in [0,\widehat{T}\rangle \\ \frac{\left(1-e^{r(t-T)}\right)nK^*}{r} & \text{for } t \in (\widehat{T},T\rangle. \end{cases}$$

Note that W for $t \in [0,\widehat{T}\rangle$ represents discounted global welfare *during* an IEA, while W for $t \in (\widehat{T},T\rangle$ represents discounted global welfare *after* an IEA (which can be thought of as the legacy of the agreement).

The impact of an increase in ϕ on W is given by

$$\frac{\partial W}{\partial \phi} = \begin{cases} \frac{\left(1-e^{r(t-T)}\right)n}{r}\frac{\partial K^*}{\partial \phi} + \frac{e^{r(t-T)}\{cr+n[r(t-T)-1]\}+n-cr}{r^2}\frac{\partial m^*}{\partial \phi} & \text{for } t \in [0,\widehat{T}\rangle \\ \frac{\left(1-e^{r(t-T)}\right)n}{r}\left(\frac{\partial K^*}{\partial \phi} + \frac{\partial K^*}{\partial \widehat{T}}\frac{\partial \widehat{T}}{\partial \phi}\right) & \text{for } t \in (\widehat{T},T\rangle. \end{cases} \tag{E.1}$$

Let $d\phi > 0$ be sufficiently large (i.e. such that m^* decreases). We know from Corollaries 2 and 4 that an increase in ϕ leads to an increase in $\widehat{T}$ and to a decrease in m^*, respectively. Moreover,

$$\frac{\partial K^*}{\partial \phi} = \frac{t\,(1-cr) + c\left[\log\left(1-e^{r(t-T)}\right) - \log\left(1-e^{-rT}\right)\right]}{\phi^2} < 0,$$

since $c > \underline{c}$ by Assumption A2. Furthermore, $\partial K^*/\partial \widehat{T} > 0$, since K^* is monotonically increasing in t. Therefore, *a priori*, the impact of ϕ on W is ambiguous.

It can be established that, for $T \to \infty$, W is decreasing in ϕ, since

$$\lim_{T\to\infty} \frac{\partial W}{\partial \phi} = \frac{n}{r}\frac{\partial K^*}{\partial \phi} + \frac{n-cr}{r^2}\frac{\partial m^*}{\partial \phi},$$

which is negative given that $\partial K^*/\partial \phi < 0$, $\partial m^*/\partial \phi < 0$, and $c < n/r$ by Assumption A2.

For $t \in [0,\widehat{T}\rangle$, from (E.1), we have

$$\begin{aligned} \frac{\partial W}{\partial \phi} \quad = \quad & \frac{e^{-2r(t-T)}\left[n\left(cr^2t-rT-1\right)+cr\right]+e^{r(t-T)}+(cr(-2+cr)+n(2+r(t+T-c(1+r(t+T)))))}{r^2\phi^2\left(e^{r(t-T)}-1\right)} \\ & - \frac{cnr\left(e^{r(t-T)}-1\right)^2\left[\log\left(1-e^{r(t-T)}\right)-\log\left(1-e^{-rT}\right)\right]+(cr-1)[n(1+rt)-cr]}{r^2\phi^2\left(e^{r(t-T)}-1\right)}. \end{aligned} \tag{E.2}$$

(E.2) is quadratic in c, with the coefficient of c^2 equal to $1/\phi^2$. Let c_1 and c_2 be the two solutions of $\partial W/\partial \phi = 0$, with $c_1 > c_2$. It can be checked that $c_2 < \underline{c}$ and $c_1 > \widehat{c}$, implying that $\partial W/\partial \phi < 0$ for $t \in [0,\widehat{T}\rangle$.

For $t \in (\widehat{T},T\rangle$, $\partial W/\partial \phi$ can be either positive or negative, depending on the parameter values. We have verified numerically that $\partial W/\partial \phi < 0$ for c sufficiently small and T sufficiently large.

References

[1] Açıkgöz, O.T., Benchekroun, H., 2017. Anticipated international environmental agreements. *European Economic Review* 92, 306-336.

[2] Aldy, J. E., Barrett, S., Stavins, R.N., 2003. Thirteen plus one: a comparison of global climate policy architectures. *Climate Policy* 3, 373-397.

[3] Andreoni, J., 1988. Privately provided public goods in a large economy: The limits of altruism. *Journal of Public Economics* 35, 57-73.

[4] d'Aspremont, C., Jacquemin, J., Gabszewicz, J. Weymark, J., 1983. On the stability of collusive price leadership. *Canadian Journal of Economics* 16, 17-25.

[5] Barrett, S., 1994. Self-enforcing international environmental agreements. *Oxford Economics Papers* 46, 878-894.

[6] Barrett, S., 1999. A theory of full international cooperation. *Journal of Theoretical Politics* 11, 519-541.

[7] Barrett, S., 2002. Consensus treaties. *Journal of Institutional and Theoretical Economics* 158, 529-547.

[8] Battaglini, M., Harstad, B., 2020. The political economy of weak treaties. *Journal of Political Economy* 128, 544-590.

[9] Battaglini, M., Harstad, B., 2016. Participation and duration of environmental agreements. *Journal of Political Economy* 124, 160-204.

[10] Battaglini, M., Nunnari, S., Palfrey, T.R., 2014. Dynamic free riding with irreversible investments. *American Economic Review* 104, 2858-2871.

[11] Benchekroun, H., Chaudhuri, A.R., 2015. Cleaner technologies and the stability of international environmental agreements. *Journal of Public Economic Theory* 17, 887-915.

[12] Benchekroun, H., Long, N.V., 2008. The build-up of cooperative behavior among non-cooperative selfish agents. *Journal of Economic Behavior and Organization* 67, 239-252.

[13] Bergstrom, T.C., Blume, L.E., Varian, H.R., 1986. On the private provision of public goods. *Journal of Public Economics* 29, 25-49.

[14] Bernheim, B Douglas, 1986. On the voluntary and involuntary provision of public goods. *American Economic Review* 76, 789-793.

[15] Bowen, R., Georgiadis, G., Lambert, N.S., 2019. Collective choice in dynamic public good provision. *American Economic Journal: Microeconomics* 11, 243-298.

[16] Buchholz, W., Cornes, R., Rubbelke, D., 2014. Potentially harmful international cooperation on global public good provision. *Economica* 81, 205-223.

[17] Carraro, C., Siniscalco, D., 1993. Strategies for the international protection of the environment. *Journal of Public Economics* 52, 309-28.

[18] Chamberlin, J., 1974. Provision of collective goods as a function of group size. *American Political Science Review* 65, 707-716.

[19] Colombo, L., Labrecciosa, P., 2018. Consumer surplus-enhancing cooperation in a natural resource oligopoly. *Journal of Environmental Economics and Management* 92, 185-193.

[20] Cyert, R.M., deGroot, M.H., 1973. An analysis of cooperation and learning in a duopoly context. *American Economic Review* 63, 24–37.

[21] Cornes, R., Sandler, T., 1986. The theory of externalities, public goods and club goods, Second Edition. Cambridge, U. K.: Cambridge University Press.

[22] de Zeeuw, A., 2008. Dynamic effects on the stability of international environmental agreements. *Journal of Environmental Economics and Management* 55, 163-174.

[23] Diamantoudi, E., 2005. Stable cartels revisited. *Economic Theory* 26, 907-921.

[24] Diamantoudi, E., Sartzetakis, E.S., 2015. International environmental agreements: coordinated action under foresight. *Economic Theory* 59, 527-546.

[25] Dutta, P.K., Radner, R., 2004. Self-enforcing climate-change treaties. *Proceedings of the National Academy of Sciences* 101, 4746-4751.

[26] Edgeworth, F.Y., 1881. Mathematical psychics: an essay on the application of mathematics to the moral sciences. Kegan Paul, London.

[27] Eichner, T., Pethig, R., 2011. Carbon leakage, the green paradox, and perfect future markets. *International Economic Review* 52, 767-805.

[28] Eichner, T., Pethig, R., 2013. Self-enforcing environmental agreements and international trade. *Journal of Public Economics* 102, 37-50.

[29] Fershtman, C., Nitzan, S., 1991. Dynamic voluntary provision of public goods. *European Economic Review* 35, 1057-1067.

[30] Finus, M., Maus, S., 2008. Modesty may pay!. *Journal of Public Economic Theory* 10, 801-826.

[31] Fujiwara, K., Matsueda, N., 2009. Dynamic voluntary provision of public goods: An extension. *Journal of Public Economic Theory* 11, 27-36.

[32] Georgiadis, G., 2015. Project and Team Dynamics. *Review of Economic Studies* 82, 187-218.

[33] Georgiadis, G., 2017. Deadlines and infrequent monitoring in the dynamic provision of public goods. *Journal of Public Economics* 152, 1-12.

[34] Gerlagh, R., Liski, M., 2018. Consistent climate policies. *Journal of the European Economic Association* 16, 1-44.

[35] Hagen, A., Schneider, J., 2021. Trade sanctions and the stability of climate coalitions. *Journal of Environmental Economics and Management* 109, 102504.

[36] Harstad, B., 2020a. Pledge-and-review bargaining: A theory. Mimeo.

[37] Harstad, B., 2020b. Pledge-and-review bargaining: From Kyoto to Paris. Mimeo.

[38] Harstad, B., 2020c. Technology and time inconsistency. *Journal of Political Economy* 128, 2653-2689.

[39] Hoel, M., 1992. International environmental conventions: the case of uniform reductions of emissions. *Environmental and Resource Economics* 2, 141-159.

[40] Hoel, M., de Zeeuw, A., 2014. Technology agreements with heterogeneous countries. CESifo Working Paper No. 4635.

[41] Hong, F., Karp, L., 2012. International environmental agreements with mixed strategies and investment. *Journal of Public Economics* 96, 685-697.

[42] Itaya, J., Shimomura, K., 2001. A dynamic conjectural variations model in the private provision of public goods: a differential game approach. *Journal of Public Economics* 81, 153-172.

[43] Kalai, E., 1977. Proportional solutions to bargaining situations: Intertemporal utility comparisons. *Econometrica* 45, 1623-1630.

[44] Kalai, E., Smorodinsky, M., 1975. Other solutions to Nash's bargaining problem. *Econometrica* 43, 513-518.

[45] Karp, L., Sakamoto, H., 2021. Sober optimism and the formation of international environmental agreements. *Journal of Economic Theory* 197, 105321.

[46] Karp, L., Simon, L.K., 2013. Participation games and international environmental agreements: A non-parametric model. *Journal of Environmental Economics and Management* 65, 326-344.

[47] Khourdajie, A.A., Finus, M., 2020. Measures to enhance the effectiveness of international climate agreements: The case of border carbon adjustments. *European Economic Review* 24, 103405.

[48] Köke, S., Lange, A., 2017. Negotiating environmental agreements under ratification constraints. *Journal of Environmental Economics and Management* 83, 90-106.

[49] Kolstad, C. D., Toman, M., 2005. The economics of climate policy. Handbook of Environmental Economics 3, 1562-93.

[50] Kolstad, C.D., 2011. Environmental Economics, second edn. Oxford University Press, New York.

[51] Kovác, E., Schmidt, R.C., 2021. A simple dynamic climate cooperation model. *Journal of Public Economics* 194, 104329.

[52] Lopez, A., Vives, X., 2019. Overlapping ownership, R&D spillovers and antitrust policy. *Journal of Political Economy* 127, 2394-2437.

[53] Mukanjari, S., Sterner, T., 2018. Do markets trump politics? Evidence from fossil market reactions to the Paris Agreement and the U.S. election. Working Paper in Economics, No. 728, University of Gothenburg, Gothernburg.

[54] Marx, L.M., Matthews, S.A., 2000. Dynamic voluntary contribution to a public project. *Review of Economic Studies* 67, 327-358.

[55] Morrison, C.C., 1978. A note on providing public goods through voluntary contributions. *Public Choice* 33, 119-123.

[56] Nash, J., 1950. The bargaining problem. *Econometrica* 18, 155-162.

[57] Nordhaus, W.D., 2015. Climate clubs: Overcoming free-riding in international climate policy. *American Economic Review* 105, 1339-1370.

[58] Rubio, S.J., Ulph, A., 2006. Self-enforcing international environmental agreements revisited. *Oxford Economic Papers* 58, 233-263.

[59] Rubio, S.J., Ulph, A., 2007. An infinite-horizon model of dynamic membership of international environmental agreements. *Journal of Environmental Economics and Management* 54, 296-310.

[60] Schmalensee, R., 1998. Greenhouse policy architectures and institutions. Economics and Policy Issues in Climate Change, ed. by W. D. Nordhaus, Resources for the Future Press, Washington, D. C.

[61] Ulph, A., 2004. Stable International Environmental Agreements with a stock pollutant, uncertainty and learning. *Journal of Risk and Uncertainty* 29, 53-73.

[62] van der Ploeg, R., Rezai, A., 2020. The risk of policy tipping and stranded carbon assets. *Journal of Environmental Economics and Management* 100, 102258.

[63] van der Pol, T., Weikard, H. P., & van Ierland, E., 2012. Can altruism stabilise international climate agreements?. *Ecological Economics* 81, 112-120.

[64] Varian, H. R. 1994. Sequential contributions to public goods. *Journal of Public Economics* 53, 165-186.

[65] Wirl, F., 1996. Dynamic voluntary provision of public goods: Extension to nonlinear strategies. *European Journal of Political Economy* 12, 555-560.

[66] Yanase, A., 2006. Dynamic voluntary provision of public goods and optimal steady-state subsidies. *Journal of Public Economic Theory* 8, 171-179.