Ecological Economics 211 (2023) 107869

Contents lists available at ScienceDirect

Ecological Economics

journal homepage: www.elsevier.com/locate/ecolecon

Self-image and the stability of international environmental agreements

Michèle Breton [a], Lucia Sbragia [b,*]

[a] HEC Montréal, 3000 Chemin de la Côte Sainte-Catherine, Montréal, QC H2T 2A7, Canada
[b] Durham University Business School, Millhill Ln, DH1 3LB, Durham, UK

ARTICLE INFO

Keywords:
Dynamic game
Climate change
Stability

ABSTRACT

In this paper we examine the stability of international environmental agreements about a (common) emissions target. By signing the agreement, the parties develop a sense of responsibility to the commitment made, gaining a self-image that contributes to their utility.

We study a dynamic two-stage game where all countries act individualistically. We investigate how two fundamental components of the model, that is, the ambition of the pledge and the relative importance given to compliance to the commitment, affect the stability and efficiency of the agreement in terms of global welfare and total emissions.

We find that participation is the key driver of all the results and that it is negatively related to the ambition of the pledge and positively related to countries' level of concern about environmental issues.

1. Introduction

During the last two years, several individual national states, countries, territories, and councils published declarations about climate emergency (for instance, the UK in May 2019, Canada and France in June 2019, the EU in November 2019, Japan in November 2020, Singapore in February 2021), indicating that climate change is an important issue in the political agenda of countries worldwide.

In the long history of international cooperation to tackle the problem of climate change, which started in 1988 with the establishment of the Intergovernmental Panel on Climate Change (IPCC), the Paris Agreement (2015) can be considered the last milestone. In short, each country that joins the Agreement publicly declares its level of ambition in the race to keep the temperature increase within the agreed range.[1] These pledges are not binding, meaning that no sanctions are imposed on members who fail to meet their promise. However, the periodic review process creates regular moments for *naming and shaming* strategies to be played against those countries that do not meet their international commitments. The outcome of the Paris Agreement has left experts with mixed feeling: on one hand, it was able to achieve a large participation of countries (195 Parties have signed and 189 have ratified the Agreement);[2] on the other hand, the pledges to reduce emissions made by countries at Paris seemed insufficient to meet the agreed temperature target.[3]

In this paper we contribute to the literature on International Environmental Agreements (IEAs) by proposing a stylized model where international cooperation is promoted by a mix of "naming and shaming" (similar to the logic of the Paris Agreement) and "warm glow" (Andreoni, 1990). Using this model, we examine how the level of ambition and environmental concern of signatory countries affect the stability and the efficiency of an agreement in terms of global welfare and total emissions.

* Corresponding author.
E-mail addresses: michele.breton@hec.ca (M. Breton), lucia.sbragia@durham.ac.uk (L. Sbragia).

1 "Holding the increase in the global average temperature to well below 2 degrees Celsius above pre-industrial levels and pursuing efforts to limit the temperature increase to 1.5 degrees Celsius above pre-industrial levels" (Article 2).

2 https://www.un.org/en/climatechange/paris-agreement#:~:text=Entered%20into%20force%20less%20than,have%20joined%20the%20Paris%20Agreement

3 Climate Interactive projected the national plans would curb temperature rises to 3.5 degrees Celsius, compared with 4.5 degrees Celsius if no action was taken. Climate Action Tracker projected the plans, if implemented, would limit average temperature rises to 2.7 degrees Celsius above pre-industrial times by 2100.

4 In this paper, we assume that the target has already being decided, for instance as the result of an initial agreement, or by an international panel of experts.

5 National and international environmental campaign groups, media, common citizens concerned about environmental problems, where pressures from civil society can be reinforced by international criticisms on missed compliance.

6 The use of naming and shaming as a strategy to create social support to enforce an IEA has been empirically tested by Barrett and Dannenberg (2016) and Tingley and Tomz (2021); while experiments about a warm glow coming from a proenvironmental behavior include Khaneman and Knetsch (1992), Hartmann et al. (2017) and Venhoeven et al. (2016).

https://doi.org/10.1016/j.ecolecon.2023.107869
Received 8 April 2022; Received in revised form 25 April 2023; Accepted 26 April 2023
Available online 29 May 2023

Specifically, we develop a multi-stage dynamic game over an infinite horizon, where the state variable is the level of accumulated pollution. Symmetric players first decide on whether or not to participate in an IEA (*membership game*), and then decide about their emissions levels as a function of the current level of the pollution stock (*emission game*). Countries with different participation status (signatories vs non signatories) show heterogeneous preferences in the emission game.

We assume that the IEA is about endorsing some common target, whose value characterizes the ambition of the agreement.[4] When a country publicly becomes a signatory of the IEA, it develops a sense of responsibility towards the commitment taken, a *self-image*, as its actions are then measured against the target, at both the international and national levels. In particular, we assume that the utility of a signatory country includes a self-image component, related to the commitment taken. The self-image is comparable to a performance indicator, accounting for the outcome of a naming and shaming strategy deployed by both peer states and the civil society[5] (Falkner, 2016), as well as for a warm glow from participating in the effort to curb global warming. Accordingly, a signatory country that is able to overcomply with respect to the common target experiences a positive utility (sense of pride, warm glow), while not respecting its commitment creates a negative utility (sense of shame).[6] We further assume that the relative weight of the self-image in the utility function of signatory countries depends on the seriousness of the global environmental problem, on countries awareness and concern about it, and on the ambition of the signatories' commitment.

We first investigate whether this setting can generate stable agreements of meaningful size and impact. We then move our attention to the analysis of the relationship between the ambition of the pledge and the size of a stable IEA, and its environmental and economic consequences. Finally, motivated by the increasing number of initiatives, in person and on various media, to inform and educate people about climate change and the risks associated with it, we investigate how the environmental concern of countries affects the stability of an IEA.

Our first contribution is to show that, when countries care about their self-image, either as a shame or as a warm glow, it is possible to design stable agreements that have a positive impact on both global welfare and the environment; that is, both signatory and nonsignatory countries are better off, and the steady state pollution stock is lower, than in the no-agreement solution. This responds to the conclusions reached by Barrett and Dannenberg (2016) where it is stated that the pledge and review mechanism should be combined with other measures. We also find that there is a trade-off between ambition and participation in an IEA, with participation having a leading effect on the environment and global welfare. Less ambitious targets stimulate greater participation in the IEA, and this, in turn, reduces the global stock of pollution, allowing for greater global welfare. This result is robust to the model's parameter values, as well as when we assume that the agreed upon target adapts to the global pollution level.

Our paper is related to two streams of literature. The first is the literature on the stability of IEAs, under the so-called *noncooperative approach*. Starting from the founding works by Hoel (1992), Carraro and Siniscalco (1993), and Barrett (1994), this literature has developed abundantly over the years; see for instance, to mention only a few, Petrakis and Xepapadeas (1996), Botteon and Carraro (1997), Hoel and Schneider (1997), Barrett (1997), Carraro and Siniscalco (1997), and Carraro et al. (2009) in a static framework; Rubio and Ulph (2007), de Zeeuw (2008), and Breton et al. (2010) in a dynamic context; and, more recently, Finus and Rübbelke (2013), Weikard et al. (2015), and Marchiori et al. (2017). Within this literature, some authors have considered the possibility that countries' objectives include "other-regarding" considerations: in Lange and Vogt (2003) and Lange (2006), countries share a sense of equity; in Grüning and Peters (2010), countries' preferences incorporate justice and fairness; and in van der Pol and van Ierland (2012), altruism affects all countries, albeit in different ways. A fundamental difference between all these contributions and our model is in the nature of the agreement. In the standard literature on IEAs, signatory countries agree to coordinate their actions, while in our paper, signatory countries instead agree on a target, similarly to the Paris Agreement, leaving them the freedom to reach it in the way most convenient to them. This has two consequences: the first is that all countries, including signatory countries, act non-cooperatively in the emissions game. A second difference is that, since nonsignatory countries do not make any commitment, their utility does not include any self-image concerns, providing a different kind of asymmetry to the emissions game.

The second stream is the self-image literature. We identify this literature with a specific instance of social preferences models, which can be classified into three groups: models with status-concerned players, where players compare their income (consumption, utility) to those of their peers (as in Frank, 2005; Card et al., 2012, and Benchekroun and Van Long, 2016; models with altruistic players, who include other players' welfare in their objective function (as in Fehr and Schmidt, 1999 and Colombo and Labrecciosa, 2018); and finally self-image models, where players are morally concerned about a given norm and develop a self-image with respect to it, that contributes to their utility function. Contributions within this literature usually consider a single type of agent, labeled *socially responsible*, whose utility includes both monetary and moral benefits derived from doing the "right thing." For instance, in Brekke et al. (2003), socially responsible consumers strategically decide about allocating their time between contributing to a public good and leisure, and their utility function includes a penalty when their contribution does not reach the morally ideal effort; in Eyckmans and Kverndokk (2010), morally concerned countries manage a pollution-permit trading scheme; in Van Long (2020, 2021), the utility function of agents exploiting a common-property renewable resource (a fishery) is affected by a sense of shame induced by violating the norm.

The paper that is closest to our contribution is Wirl (2011), which considers two types of players, namely, "brown" and "green" countries, in an environmental context. As in the self-image literature, the countries' type is based on differing preferences. Green countries suffer, in addition to the environmental damage, an increasing and convex penalty function for emitting more than the norm, which is given by the cooperative (first-best) solution. A differential emissions game is solved for a given composition of brown and green countries, where both types of countries choose their level of emissions by maximizing their individual welfare. Our paper differs from Wirl (2011) in two main regards. Firstly, from a modeling point of view, we adopt a self-image that can generate both a sense of shame and a sense of pride, thus allowing for the possibility that the warm glow experienced by green countries for doing the right thing compensates them for their loss of welfare with respect to the free-riders; and, secondly, in terms of analysis, we are incorporating the issue of stability in the context of an environmental agreement, which allows us to determine whether a given composition, in terms of green and brown countries, can actually materialize, and under what conditions.

Our research question is related to an issue raised in Barrett (2003), that is, the possible trade-off between the breadth of international

[4] In this paper, we assume that the target has already being decided, for instance as the result of an initial agreement, or by an international panel of experts.

[5] National and international environmental campaign groups, media, common citizens concerned about environmental problems, where pressures from civil society can be reinforced by international criticisms on missed compliance.

[6] The use of naming and shaming as a strategy to create social support to enforce an IEA has been empirically tested by Barrett and Dannenberg (2016) and Tingley and Tomz (2021); while experiments about a warm glow coming from a proenvironmental behavior include Khaneman and Knetsch (1992), Hartmann et al. (2017) and Venhoeven et al. (2016).

cooperation (in terms of the number of participants) and its depth (in terms of the actions agreed upon by the parties): Is a "broad but shallow" treaty better than a "narrow but deep" one? Our findings seem to indicate that the number of participants is the driving factor towards a better outcome in the climate change context.

The rest of this paper is organized as follows. Section 2 describes the stylized model used to characterize the welfare of countries, the salient components of an IEA, and the self-image functions used in this paper. Section 3 characterizes the solution of both the membership and the emissions games among countries, where a subset of (green) countries participate in an IEA. Section 4 provides numerical illustrations, analyzing the impact of model parameters on the equilibrium solution and on the corresponding long-term global welfare and pollution stock, and proposing examples of time trajectories for the stock of pollution and the number of signatories according to various assumptions about the way countries could adhere to an IEA. Section 5 is a short conclusion. The Appendix contains detailed analytical developments and additional illustrations.

2. Model

2.1. Welfare

We consider N symmetric countries, whose production activities create economic value but also pollution emissions as a by-product. We assume that the emissions generated by a country are proportional to its production level,[7] and that the net revenue derived from country i's production activity in a given period is quadratic. Denote by $x_{it} \geq 0$ the emissions generated by the production of country i in time period t. We normalize the emissions units so that emissions are equal to 1 at the production level that maximizes revenue. Accordingly, the net revenue of country i, expressed as a function of its pollution emissions in period t, is given by

$$R_i(x_{it}) = \left(1 - \frac{1}{2}x_{it}\right)x_{it}.$$

Countries' polluting emissions accumulate over time; the time evolution of the pollution stock is assumed to be governed by the linear discrete-time equation

$$P_t = \delta P_{t-1} + \sum_{i=1}^{N} x_{it} \tag{1}$$

where $1 - \delta \in (0, 1)$ is the natural decay rate of the pollution stock.

In each period, countries suffer an environmental damage cost arising from the accumulated global pollution stock, which is assumed quadratic, increasing, and convex, so that the environmental damage suffered by each country in time period t is given by

$$D(P_t) = d_1 P_t + d_2 P_t^2,$$

with $d_1 \geq 0$ and $d_2 > 0$. As a result, the welfare of country i over an infinite horizon is given by

$$W_i = \sum_{t=0}^{\infty} \beta^t \left(\left(1 - \frac{x_{it}}{2}\right)x_{it} - d_1 P_t - d_2 P_t^2\right) \quad \text{s.t. (1)} \tag{2}$$

where $\beta \in (0, 1)$ is the periodic discount factor.

2.2. Agreement

The above assumptions characterize the standard stylized model commonly used in the IEA literature,[8] with countries divided into two groups, where a number $n_t \in [2, N]$ of countries, identified as "signatories," have agreed to participate in an IEA. In this literature, it is usually assumed that participating countries agree to coordinate their emissions levels (or, equivalently, their abatement levels), while the nonparticipating countries act noncooperatively.

In this paper, in line with recent international agreements aimed at reducing global atmospheric emissions leading to climate change, we rather assume that countries participating in an IEA agree on adjusting their polluting emissions with respect to some (common) target value, but do not coordinate their emissions decisions. We partition the set of countries into two groups, where "green" countries (indexed by G) agree to participate in the agreement and "brown" countries (indexed by B) do not, and we denote by n (*resp.* m) the number of green (*resp. brown*) countries, with $n + m = N$. We consider a game in two stages, where the first stage is the membership game, in which players decide whether or not to adhere to the agreement, and where the second stage is the emissions game, in which all players decide on their emissions level independently.

What differentiates green countries from brown countries is the fact that the utility of countries participating in the agreement consists of two components: a welfare component, described by Eq. (2), and a self-image component, related to the extent to which a country complies with the pledge making up the agreement. Brown countries, which have not pledged to attain any target, only consider their welfare (production revenues minus environmental cost) when deciding about their emissions.

2.3. Self-image

We model self-image as a linear, symmetric function of the form

$$S(x, T) = T - x,$$

where T is the target. A linear self-image can take negative or positive values, according to the position of the country's emissions with respect to the target. When emissions are above the target, the contribution of self-image to the country's utility is negative (shaming). On the other hand, when emissions are below the target, its contribution is positive (warm glow). Symmetry in a self-image function refers to the fact that the contributions of a positive or of a negative deviation with respect to the target are the same in absolute value.

Accordingly, the utility of a green country, including the self-image component, is given by

$$U_i^G = \sum_{t=0}^{\infty} \beta^t \left(\left(1 - \frac{x_{it}}{2}\right)x_{it} - d_1 P_t - d_2 P_t^2 + \lambda_t S_{it}\left(x_{it}, T\right)\right) \quad \text{s.t. (1)},$$

where λ_t is the weight of the self-image. This weight characterizes the relative importance of self-image with respect to economic welfare in the utility of green countries.

It is reasonable to assume that the relative importance of self-image in the utility of green countries is increasing with the seriousness of the global environmental problem; for instance, if the accumulated stock of pollution is very high, the weight of the self-image component is expected to be higher than when the accumulated stock of pollution is low, so that the warm glow obtained from taking action to fix the problem (or the shame from not keeping one's promise to do so) becomes relatively more important. It is also reasonable to assume that the weight of the self-image of green countries is increasing with

[7] Note that we only consider the emissions generated by the production of a generic good; emissions from consumption are assumed negligible.

[8] In most cases, either d_1 or d_2 is set to 0, giving rise to two standard models classified as linear- or quadratic-damage.

the ambition of their pledge: attaining a difficult target and making a significant reduction in one's emissions should provide more warm glow than attaining a less ambitious one, by pledging, for instance, to maintain the status quo. Finally, the weight of the self-image should also be increasing with countries' awareness of climate change and of the risks associated with it.

We use the simple specification

$$\lambda_t = \gamma \alpha P_t \tag{3}$$

to model the impact of these three factors, where the accumulated pollution is used as a proxy of the seriousness of the environmental problem, γ is the *concern* parameter and α is the *ambition* of the pledge. The concern parameter γ can be enhanced through education campaigns and information dissemination, while the ambition parameter α is a design parameter that is inversely related to the agreed upon target (since we are considering a public bad, more ambitious pledges correspond to emission targets that are lower than less ambitious ones).[9] In order to characterize the relative ambition of pledges, we compare the agreed upon target T to some arbitrary fixed reference; specifically, the target is related to the ambition parameter α by the following relation

$$T = \frac{r}{\alpha},$$

where r is a reference point representing a common norm. This can be a simple benchmark, such as the business-as-usual, the current, the pre-industrial, or the first-best emissions level.[10] We also consider the possibility of a more sophisticated reference point and corresponding target that adapt to the state of the system (here, to the accumulated stock of pollutants). Assuming that the target and reference point are linear (typically decreasing) in the pollution level, $r \equiv r_0 - r_1 P_t$, yields the following self-image function for country i in time period t

$$S_{it}\left(x_{it}, P_t\right) = \frac{r_0 - r_1 P_t}{\alpha} - x_{it}$$

where $r_1 = 0$ corresponds to the constant-target case.

The utility functions of brown and green countries are then given by

$$U_i^B = \sum_{t=0}^{\infty} \beta^t \left(\left(1 - \frac{x_{it}}{2}\right) x_{it} - d_1 P_t - d_2 P_t^2 \right) \tag{4}$$

$$\begin{aligned} U_i^G &= \sum_{t=0}^{\infty} \beta^t \left(\left(1 - \frac{x_{it}}{2}\right) x_{it} - d_1 P_t - d_2 P_t^2 + \gamma\alpha P_t \left(\frac{r_0 - r_1 P_t}{\alpha} - x_{it} \right) \right) \quad (5) \\ &= \sum_{t=0}^{\infty} \beta^t \left(\left(1 - \frac{x_{it}}{2} - \gamma\alpha P_t\right) x_{it} - P_t \left(d_1 - \gamma r_0\right) - P_t^2 \left(d_2 + \gamma r_1\right) \right) \\ &\quad \text{s.t. } (1), \end{aligned}$$

where it is apparent that the self-image resulting from participating in an IEA modifies the utility function of green countries with respect to brown countries in a way that makes their dependence to their own emissions and to the global pollution stock no longer symmetric.

[9] Note that by making the weight of the self-image dependent on the ambition, we avoid the possibility that agreements could generate a large positive utility by selecting easily attainable targets, such as the status quo.

[10] For instance, the UK Prime Minister set in law the new target to cut carbon emissions by 78%, compared to 1990 levels, by the year 2035. This pledge sets r at the 1990 emission level and the target at $0.78r$, which corresponds to an ambition level $\alpha = 1/0.78 = 1.28$

https://www.gov.uk/government/news/uk-enshrines-new-target-in-law-to-slash-emissions-by-78-by-2035

https://www.reuters.com/world/uk/uks-johnson-unveil-new-goal-emission-cuts-ahead-biden-summit-ft-2021-04-20/

3. Dynamic game setting

In this section, we develop and solve dynamic membership and emission game played by countries. We first characterize a benchmark first-best solution maximizing the joint welfare of all countries. We then solve a two-stage game, where in the first stage (membership game), countries decide whether or not to participate in the agreement, while in the second stage (emissions game), green and brown countries decide on their emission levels at discrete dates over an infinite horizon. The game is solved by backward induction.

Our assumption that emissions are proportional to the production level rules out the existence of actions that could decrease the pollution stock (apart from natural decay). Accordingly, in the solution of the countries' optimization problems, emissions are restricted to being non-negative. In the sequel, we assume that the parameter values defining the welfare functions (δ, β, d_1, d_2), the self-image functions (r_0, r_1, γ, α), and the number of countries (N, n) yield interior solutions.

3.1. First-best solution

The cooperative equilibrium, or first-best solution, corresponds to the emissions strategy that maximizes the total welfare of all countries. Since countries are symmetrical, the first-best solution is the feedback strategy $x = \left\{x_t, t = 0, \ldots, \infty\right\}$ solving

$$\max_{x \geq 0} \left\{ N \sum_{t=0}^{\infty} \beta^t W_t\left(x_t, P_t\right) \right\} \tag{6}$$

s.t.

$$P_t = \delta P_{t-1} + N x_t. \tag{7}$$

Proposition 1. *Let*

$$b_2^{fb} \equiv \frac{1 - \beta\left(\delta^2 + 2N^2 d_2\right) - \sqrt{\left(1 - \beta\left(\delta^2 + 2N^2 d_2\right)\right)^2 + 8N^2\beta d_2}}{4N^2\beta}$$

$$b_1^{fb} \equiv \frac{2N\beta b_2^{fb}\left(\delta + N d_1\right) - d_1}{1 - \beta\left(\delta + 2N^2 b_2^{fb}\right)}.$$

If the model parameter values $\left\{N, \delta, \beta, d_1, d_2\right\}$ *satisfy*

$$b_1^{fb} + 2\frac{N}{1-\delta}\delta b_2^{fb} > -\frac{1}{N\beta} \tag{8}$$

then the first-best emission strategy is linear in the current pollution stock and is given by

$$x^{fb}(P) = \frac{1 + N\beta\left(b_1^{fb} + 2P\delta b_2^{fb}\right)}{1 - 2N^2\beta b_2^{fb}} \in [0, 1].$$

The corresponding steady state is

$$P^{fb} = N\frac{1 + N\beta b_1^{bf}}{1 - \delta - 2N^2\beta b_2^{bf}}.$$

The first-best value function is quadratic, concave and decreasing in the current pollution stock and is given by

$$V^{fb}(P) = b_0^{fb} + b_1^{fb}P + b_2^{fb}P^2$$

where

$$b_0^{fb} = \frac{1 + N b_1^{fb}\beta\left(N\beta b_1^{fb} + 2\right)}{2(1-\beta)\left(1 - 2N^2\beta b_2^{fb}\right)}.$$

Proof. See Appendix 6.1. ■

3.2. Emissions game

We now solve the emissions game between n green and m brown countries over an infinite horizon, assuming that the number of signatories, the target, and the weight of the self-image are given when players decide on their emissions, and that all countries make their decisions independently. Furthermore, we assume that players are myopic with respect to the time evolution of the number of green countries.[11]

3.2.1. Reaction function

We first characterize the infinite horizon optimization problem faced by a single country (green or brown) when the strategies of the other countries are fixed.

Consider a single country, with an immediate utility given by (see Eqs. (4)–(5))

$$U_t\left(x_t, P_t\right) = \left(1 - \frac{x_t}{2}\right) x_t - d_1 P_t - d_2 P_t^2 + \theta \alpha P_t \left(\frac{r_0 - r_1 P_t}{\alpha} - x_t\right),$$

where

$$\theta = \begin{cases} \gamma \text{ if the country is green} \\ 0 \text{ if the country is brown.} \end{cases} \quad (9)$$

The optimization problem faced by this country is then

$$\max_x \left\{ \sum_{t=0}^{\infty} \beta^t U_t\left(x_t, P_t\right) \right\}$$
s.t.
$$P_t = \delta P_{t-1} + x_t + O_t$$

where O_t denotes the total emissions by all the other countries during period t. Denote by P the current level of the pollution stock. Since this optimization problem is over an infinite horizon, the optimal solution at any time depends only on P. The following proposition shows that if the emissions level of the other countries is linear in P, then the best response of a single country is also linear in P.

Proposition 2. *If the total of the emissions by all other countries is non-negative and linear in the pollution stock, the reaction value function of a single country is quadratic*

$$V(P) = b_2 P^2 + b_1 P + b_0$$

and the optimal strategy of this country is linear in the pollution stock, provided that $x(P) \geq 0$ *for* $P \in \left[0, \frac{N}{1-\delta}\right]$ *and that there exist constants* b_0, b_1, *and* b_2 *satisfying*

$$b_2 = -\theta r_1 - d_2 + \frac{1}{2}\theta^2\alpha^2 - \beta M^2 b_2\left(2\beta b_2 - 1\right) < 0 \quad (10)$$

$$b_1 = \frac{d_1 + \theta\left(\alpha - r_0\right) + 2\beta M Q b_2\left(2\beta b_2 - 1\right)}{-\beta M\left(2\beta b_2 - 1\right) - 1} < 0 \quad (11)$$

$$b_0 = \frac{1}{2}\frac{1 - \beta^2 b_1^2 - 2\beta Q\left(b_1 + Q b_2\right)\left(2\beta b_2 - 1\right)}{1 - \beta} \quad (12)$$

$$0 > 2\beta b_2 - 1 \quad (13)$$

where Q and M characterize the joint emission strategy from all countries,

$$X(P) = Q + (M - \delta) P.$$

Proof. See Appendix 6.2. ■

Note that the constants Q and M are global values that need to be defined according to the equilibrium solution concept, which is done in the following subsection.

[11] See Breton and Garrab (2014) for an IEA model where players account for the dynamics of the number of signatories.

As shown in the proof of Proposition 2, the reaction at P of a player to a global emission level O by the other players is

$$x(O, P) = \frac{1 + \beta b_1 + 2O\beta b_2 - P\left(\alpha\theta - 2\beta\delta b_2\right)}{1 - 2\beta b_2}.$$

Given $b_1 < 0$ and $b_2 < 0$, this means that the optimal emission of a player is decreasing with the emissions O of the other players and with the stock of pollution P – and smaller than 1, which corresponds to the production level maximizing revenues.

3.2.2. Nash equilibrium

We now assume that countries are divided into two groups of symmetric players, labeled B and G, where players in the same group have the same immediate utility, and where the number of players in group B (brown countries) is m, and n in group G (green countries), with $m + n = N$.

Proposition 2 implies that there exists a Nash equilibrium to the emissions game in non-negative linear feedback stationary strategies if we can obtain coefficients b_k^G and b_k^B, $k \in \{0, 1, 2\}$, characterizing the reaction function of each group of countries, satisfying the system (10)–(13) involving Q and M. Proposition 3 characterizes the global values Q and M corresponding to a Nash equilibrium.

Proposition 3. *If there exists a linear Nash equilibrium feedback strategy to the emissions game, the global values Q and M satisfy*

$$M = \frac{\delta - n\alpha\gamma}{1 - 2\beta\left(m b_2^B + n b_2^G\right)} \quad (14)$$

$$Q = \frac{N + \beta\left(m b_1^B + n b_1^G\right)}{1 - 2\beta\left(m b_2^B + n b_2^G\right)}. \quad (15)$$

Proof. See Appendix 6.3. ■

For a given set of parameters $\left\{N, \delta, \beta, \alpha, \gamma, d_1, d_2\right\}$ and for a given number of signatories n, conditions (10)–(15) are sufficient to determine values Q, M, and b_k^G, b_k^B, $k \in \{0, 1, 2\}$ characterizing a Nash equilibrium in linear feedback strategies. The equilibrium strategy is then given by

$$x^G(P) = 1 + \beta\left(2Q b_2^G + b_1^G\right) + P\left(2\beta M b_2^G - \alpha\gamma\right)$$
$$x^B(P) = 1 + \beta\left(2Q b_2^B + b_1^B\right) + 2P\beta M b_2^B.$$

According to Propositions 2 and 3, a Nash equilibrium in non-negative strategies exists if b_2^G, b_2^B, b_1^G and b_1^B are negative and if $x^G(P)$ and $x^B(P)$ are non-negative on $P \in \left[0, \frac{N}{1-\delta}\right]$.

The equilibrium solution corresponding to the case where all countries are green (*resp. brown*) can be obtained by setting $n = N$ (*resp.* $n = 0$). The equilibrium solution for $n = 0$ corresponds to the non-cooperative or business-as-usual equilibrium in a standard IEA model. In the sequel, we identify this special case (the *no-agreement* solution) by indexing it with na. Note that the equilibrium solution for $n = N$ does not correspond to the first-best solution, since green players do not coordinate their emission strategies.

3.2.3. Steady state

For a given n, if a Nash equilibrium in non-negative strategies exists, the steady state $P^*(n)$ of the pollution level corresponding to the equilibrium solution of the emissions game is defined by

$$\delta P_t + Q + (M - \delta) P_t = P_t,$$

yielding

$$P^*(n) = \frac{Q}{1 - M}. \quad (16)$$

3.2.4. Solution of the emission game

While analytical, the solution of the emission game cannot be obtained in closed form, which means that is not possible to provide exact conditions on the model parameters ensuring the existence of a feasible equilibrium solution to the emission game. It is however straightforward for any $n \in [0, N]$ to numerically obtain the values b_2^G, b_2^B, b_1^G and b_1^B corresponding to a given set of parameters $\{N, \delta, \beta, \alpha, \gamma, d_1, d_2\}$ and to check that the corresponding equilibrium strategies are non negative on $P \in \left[0, \frac{N}{1-\delta}\right]$. Details are provided in Appendix 6.4.

3.3. Membership game

To solve the membership game,[12] we adopt the noncooperative point of view, where successful agreements must be self-enforcing. In this context, the stability concept introduced in d'Aspremont et al. (1983), which is widely used in the IEA literature, is such that signatories have no incentive to leave the agreement, while nonsignatories have no incentive to join the agreement. This stability concept has been extended to a dynamic setting in Rubio and Ulph (2007): in a dynamic framework, players in a given group compare their total discounted utility, computed from the current state over an infinite horizon, to what they could achieve by unilaterally switching to the other group, where the utility of green and brown players is given by the equilibrium value of the emissions game.

Specifically, at a given P and for a given number n of signatories, both types of players compare their utility in or out of the agreement; therefore, the number of signatories is stable at P when

$$V^G(P;n) \geq V^B(P;n-1) \tag{17}$$

$$V^B(P;n) \geq V^G(P;n+1). \tag{18}$$

Clearly, the stability conditions, and therefore the size of a stable agreement, depend on the current pollution level. In this paper, we investigate the stability of agreements at the steady-state, that is, we look for a pair $(n^*, P^*(n^*))$ such that Eqs. (17)–(18) are satisfied at $P^*(n^*)$, where $P^*(\cdot)$ is defined in (16). We consider an open-membership situation, where players are free to join or leave the agreement at any time, so that the number of participating countries can change over time. The pair $(n^*, P^*(n^*))$ is then a steady state, in terms of both the pollution stock and the number of signatory countries.[13]

Since the solution of the emissions game, while analytic, is not in closed form, it is not possible to solve the membership game analytically. In the following section, we investigate the role of the various components of an agreement with respect to its long-term stability. Presented results are based on an extensive numerical exploration of model parameter values.

4. Results and analysis

We use the same base-case parameter values for N, δ, β, d_1, and d_2 for all our illustrative examples (see Table 1). These values are consistent with parameter values obtained using an integrated assessment model of climate change (see Bahn et al., 2009) and produce positive emissions and total welfare for the first-best and no-agreement solutions (i.e., the cooperative and the business-as-usual solutions of the standard IEA model). The results reported in the following paragraphs are qualitatively robust to the choice of the base-case parameter values.

[12] Recall that, in our model, signatories agree on a common target but do not coordinate their emissions strategies. Here, the term "membership" does not imply that players form a coalition in the emissions game.

[13] In a closed-membership situation, the membership game is played once, at $t = 0$, and the number n_0 of participating countries remains constant afterwards, so that the steady state is $(n_0, P^*(n_0))$. The stability conditions (17)–(18) will not be satisfied at $P^*(n_0)$ in general.

Table 1
Base-case parameter values.

N	δ	β	d_1	d_2
100	0.87	0.95	1.0×10^{-5}	8.0×10^{-7}

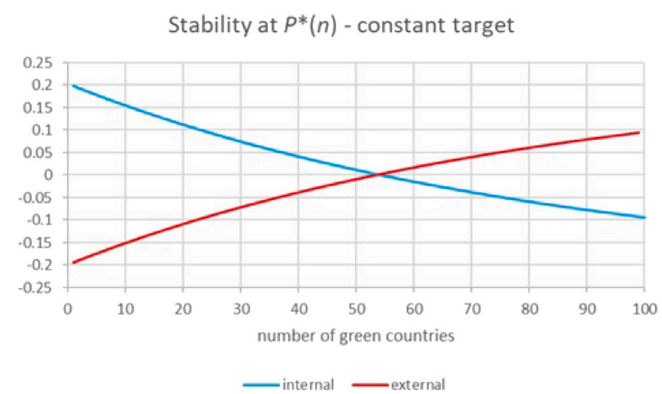

Fig. 1. Stability conditions at $P^*(n)$ as a function of the number of green (signatory) countries n. The self-image function parameter values are $r_0 = 0.594$, $\alpha = 0.7$, $\gamma = 6.5 \times 10^{-4}$. Other parameter values are listed in Table 1.

4.1. Constant target

In the first set of experiments, the reference point is a given emissions level r_0, so that the target is independent of the pollution stock and the self-image function is defined by

$$S_i(x_{it}, P_t) = \frac{r_0}{\alpha} - x_{it}.$$

Fig. 1 illustrates the solution of the membership game. To each possible number n of green countries there corresponds a steady-state pollution stock $P^*(n)$ resulting from the equilibrium solution of the emissions game and defined by Eq. (16). Fig. 1 plots the value of

(i) $V^G(P^*(n);n) - V^B(P^*(n);n-1)$ (internal stability),

(ii) $V^B(P^*(n);n) - V^G(P^*(n);n+1)$ (external stability),

as a function of n. For this specific example, the reference point is the emissions level corresponding to the first-best solution at the steady state ($x^{fb} = 0.594$ at $P^{*fb} = 457$), and the ambition parameter ($\alpha = 0.7$) results in a target $T = 0.849$, which is almost halfway between the ideal level and the no-agreement solution ($x^{na} = 0.993$ at $P^{*na} = 764$).

For this set of parameter values, both internal and external stability conditions are met at $n = 54$, and the agreement is stable at the steady state $P^*(54) = 644$. For $n < 54$, the external stability condition is violated at $P^*(n)$, and a brown country would be better off joining the agreement and increasing the number of participating countries; for $n > 54$, the internal stability condition is not met at $P^*(n)$, and a green country would find it profitable to leave the agreement.

Fig. 2 plots the equilibrium value functions (left panel) and the emissions (right panel) as a function of the current pollution stock, in the first-best solution, the no-agreement solution ($n = 0$), and the equilibrium solution between green and brown countries at $n = 54$. It also plots in the left panel the weighted average of countries' infinite-horizon welfare (TW/N), that is, not including the self-image component of the green countries' utility.

In this example, when there are 54 signatory countries, the long-term equilibrium utilities of green and brown countries are close to each other for any level of pollution stock and are situated between the no-agreement and the first-best solutions; this is also the case for the long-term average welfare of green and brown countries, as well

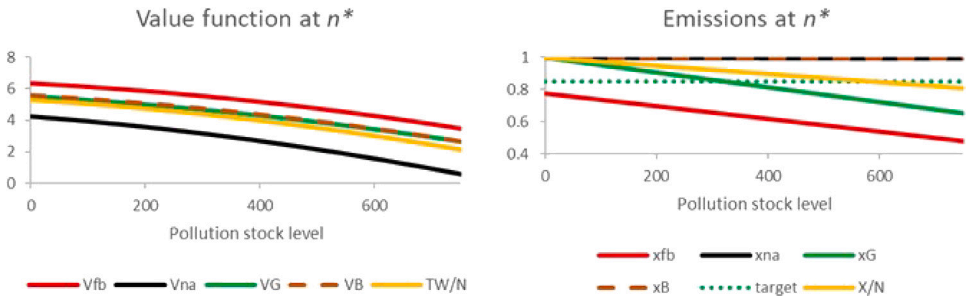

Fig. 2. Equilibrium value function (left panel) and emissions (right panel) of green and brown countries as a function of the pollution stock for $n = 54$, compared to the first-best and the no-agreement solutions. Self-image parameter values are $r_0 = 0.594$, $\alpha = 0.7$, and $\gamma = 6.5 \times 10^{-4}$. Other parameter values are from Table 1.

as the welfare of green countries.[14] Emissions of green countries are generally lower, and those of brown countries are slightly higher, than in the no-agreement case, except for very low values of the pollution stock. Furthermore, the equilibrium solution of the emissions game is such that the emissions strategy of green countries is significantly more sensitive to the current stock of pollution than that of brown countries. Given the (constant) value of the target, the self-image of green countries is negative for low values of the pollution stock ($P < 308$), as green countries emit more than the target and experience some shame for not complying with the agreement. At the steady state, the self-image of green countries is positive, as they emit less than the target, and this warm glow compensates them for the reduction in their net revenue, making their total utility comparable to that of the brown countries.

Figs. 3 and 4 illustrate an example where the self-image portion in the utility of green players is more important, as both the ambition (α) and the concern (γ) parameters take large values.[15] In this example, the ambition parameter ($\alpha = 0.837$) produces a more stringent target $T = 0.71$ and the concern parameter ($\gamma = 0.0013$) is twice as high as in Figs. 1–2. For this set of parameter values, the agreement is stable for the same number of signatory countries ($n^* = 54$) as in the previous example, but it results in a lower steady-state pollution stock ($P^*(54) = 528$). At the steady state, the green countries emit less than the target, and even less than in the first-best solution, while the brown countries emit slightly more than in the no-agreement case. With respect to the previous example, the utility of both types of countries is higher at all P. As in the preceding example, the weighted average infinite-horizon welfare is situated between the no-agreement and the first-best solutions; actually, the infinite-horizon welfare of green countries (not depicted in the figure) is slightly higher than in the no-agreement case, except for very high values of the pollution stock where it is slightly lower.

From these two examples, we can gather that stable, self-enforcing agreements can be obtained when the utilities of signatory countries include a self-image component related to their commitment to the agreement. Moreover, such agreements are effective, as they can reduce the global pollution stock and increase the global welfare with respect to the no-agreement solution, even though signatory countries act noncooperatively.

Knowing that a stable IEA can be achieved, and drawing inspiration from the current debate and events related to climate change, we now investigate the impact of the ambition of the pledge (α) and the concern level (γ) on the effectiveness of an IEA when the target is a given constant. The ambition of the pledge is a design parameter, which could be the result of discussions among an initial group of countries. The concern level depends on a country's population's awareness of climate change and the risks associated to it, and could be enhanced through education campaigns and information dissemination.

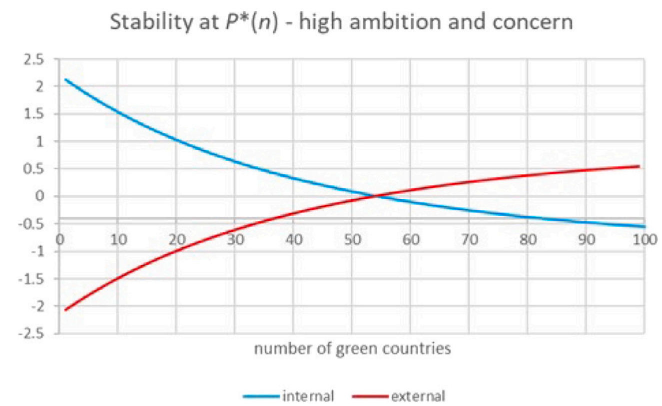

Fig. 3. Stability conditions at $P^*(n)$ as a function of the number of green (signatory) countries n. The self-image function parameter values are $r_0 = 0.594$, $\alpha = 0.837$, $\gamma = 13 \times 10^{-4}$. Other parameter values are listed in Table 1.

4.1.1. Impact of the ambition parameter

Figs. 5 and 6 illustrate the impact of the ambition of the pledge on the equilibrium solution of the membership and emissions games when the target is constant. These results are obtained by unilaterally varying the value of α in the setting described in Fig. 1. An increase in the value of α has two opposite effects: it decreases the target, making compliance more expensive, and increases the weight of self-image and therefore the value of compliance.

Fig. 5 shows the impact of the ambition parameter on the solution of the membership game and the corresponding steady-state pollution stock. When α is sufficiently high ($\alpha \geq \overline{\alpha}$), no country will join the agreement and α has no impact on the solution, while full participation is attained when α is sufficiently low ($\alpha \leq \underline{\alpha}$).[16] For $\alpha \in (\underline{\alpha}, \overline{\alpha})$, the number of participating countries at equilibrium is decreasing with the ambition parameter, while the equilibrium steady-state pollution stock is increasing with α. This illustrates that there is a negative relationship between ambition and participation, which has consequences on environmental quality.

The left panel of Fig. 6 shows the impact of α on the individual emissions of green and brown countries. Emissions by green countries are always significantly lower than emissions by brown countries. For $\alpha \in (\underline{\alpha}, \overline{\alpha})$, the emissions in both groups are decreasing with α, that is,

[14] Details on the computation of the long-term welfare of green countries are provided in Appendix 6.5.

[15] Note however that not all combinations of the self-image function parameters will give rise to a stable agreement at the steady state. When the weight of self-image is too high, the equilibrium emissions from green countries may become negative, whereas when it is too low, there may not exist a stable agreement for a target smaller than 1.

[16] Note that, under full participation, changes in $\alpha \in (0, \underline{\alpha})$ keep affecting the utility obtained by green countries.

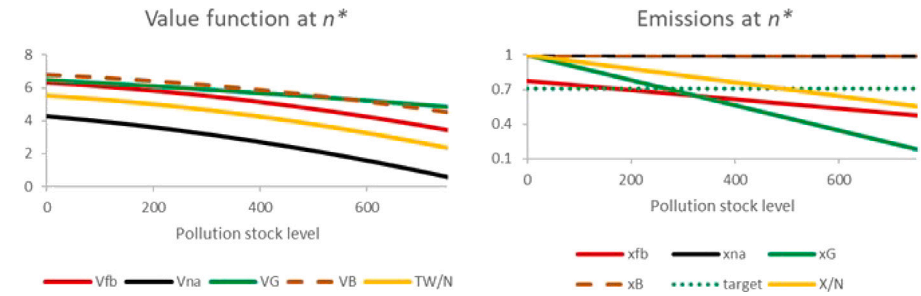

Fig. 4. Equilibrium value function (left panel) and emissions (right panel) of green and brown countries as a function of the pollution stock for $n = 54$, compared to the first-best and the no-agreement solutions. Self-image parameter values are $r_0 = 0.594$, $\alpha = 0.837$, $\gamma = 13 \times 10^{-4}$. Other parameter values are from Table 1.

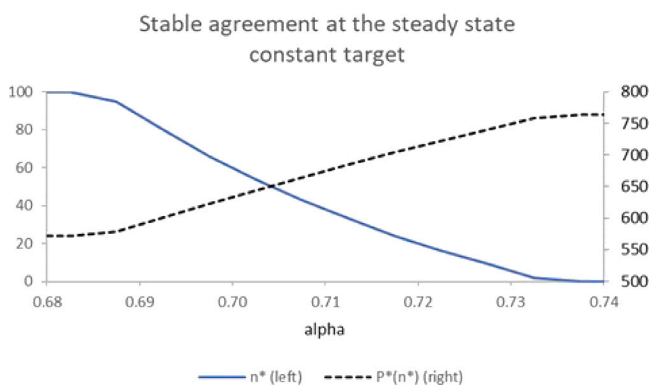

Fig. 5. Left axis: number of green countries in a stable agreement at the steady state. Right axis: pollution stock at the steady state for a stable agreement. Values are represented as a function of the ambition parameter α for $r_0 = 0.594$ and $\gamma = 0.00065$. Other parameter values are listed in Table 1.

the emissions of brown countries decrease with the ambition of green countries, but at a much lower rate than that of green countries. The decrease in emissions by both types of countries can be explained by the fact that all countries' emissions are decreasing in the pollution stock, which is increasing with α. However, green countries have an additional incentive to reduce their emissions with the value of the ambition parameter, that is, compliance with the agreement. Note that for $\alpha < \underline{\alpha}$ (full participation), the impact of α on the emissions from (green) countries is negative (decreasing α increases emissions), but much smaller than for $\alpha \in (\underline{\alpha}, \overline{\alpha})$, since α no longer impacts participation.

It is interesting to note that, even though the emissions levels for both types of countries are decreasing with the ambition of the pledge, the total emissions are increasing with α, as illustrated in the right panel of Fig. 6. This result is driven by the increase in the proportion of brown countries, which have higher emissions levels than green countries. From an environmental point of view, an agreement with a modest target seems to perform better than an agreement to attain an ambitious one, achieving a higher participation, which results in an overall decrease of total emissions.

Finally, Fig. 7 shows the impact of the ambition parameter on global welfare. Under a stable agreement, the total discounted utilities of green and brown countries are close, satisfying stability conditions (17)–(18). Since both types of countries suffer the same environmental damage cost and green countries emit less, this means that the revenue of green countries is lower than that of brown countries, but this is compensated for by the positive utility they derive from their self-image. Fig. 7 shows that both the total discounted equilibrium welfare and the total discounted equilibrium utility at the steady state are decreasing with the ambition parameter α. This figure also shows that the total self-image component makes up a relatively low proportion of the total utility of countries, even when the number of green countries is high.

Our numerical experiments with various combinations of parameter values indicate that results are qualitatively robust in the range where it is possible to obtain a stable agreement with feasible emissions levels; in equilibrium, the number of green countries, the individual emissions, the total utility, and the total welfare are decreasing with α, while the total emissions and pollution stock are increasing with α. The overall conclusion of these experiments is that, in the constant-target context, agreements that have modest targets are more efficient, achieving a higher participation, lower pollution stock, and higher global welfare.

4.1.2. Impact of the concern parameter

We now perform a similar sensitivity analysis by unilaterally varying the concern parameter γ, again in the setting described in Fig. 1. Increasing γ increases the weight of self-image and therefore has a direct impact on the value of compliance. Results are reported in Figs. 8 to 10.

Fig. 8 shows the impact of the level of the concern parameter γ on the solution of the membership game and on the steady-state pollution stock. For sufficiently small γ ($\gamma \leq \underline{\gamma}$), there is no incentive for countries to participate in the agreement, while for $\gamma \geq \overline{\gamma}$, full participation is achieved. For $\gamma \in \left(\underline{\gamma}, \overline{\gamma}\right)$, the number of participating countries is increasing with γ, which results in a decrease in the pollution stock at the steady state. Not surprisingly, increasing the weight of self-image, for instance by increasing the population's awareness of the risks of climate change, has a positive impact on the size of a stable agreement and on the quality of the environment.

The left panel of Fig. 9 shows the impact of γ on the individual emissions of green and brown countries. The emissions of green countries are lower than those of brown countries for any level of γ. For $\gamma \leq \underline{\gamma}$, the equilibrium solution of the emissions game coincides with the no-agreement solution ($n = 0$) and does not depend on γ. For $\gamma \in \left(\underline{\gamma}, \overline{\gamma}\right)$, the emissions of brown countries are increasing and the emissions of green countries are decreasing with γ. Brown countries' emissions increase with the environmental awareness of green countries because they respond to the stock of pollution, which is decreasing in γ. In the case of green countries, two opposing effects are at work when γ increases: as for brown countries, green countries increase their emissions in response to a lower stock of pollution; however, an increase in γ gives more weight to the distance between their action and the target in the green countries' utility, resulting in a decrease in their emissions. Recall that the weight of self-image is proportional to P. In all our numerical investigations, the weight $\alpha\gamma P^*(n^*)$ of self-image at equilibrium is increasing with γ, and its impact dominates that of the reduction in the pollution stock, so that green countries' emissions are decreasing with γ. When full participation is reached, the impact of γ on the pollution stock is no longer significant since the reduction in the pollution stock is driven by the number of green countries, and the negative effect of the increase in the self-image's weight becomes predominant.

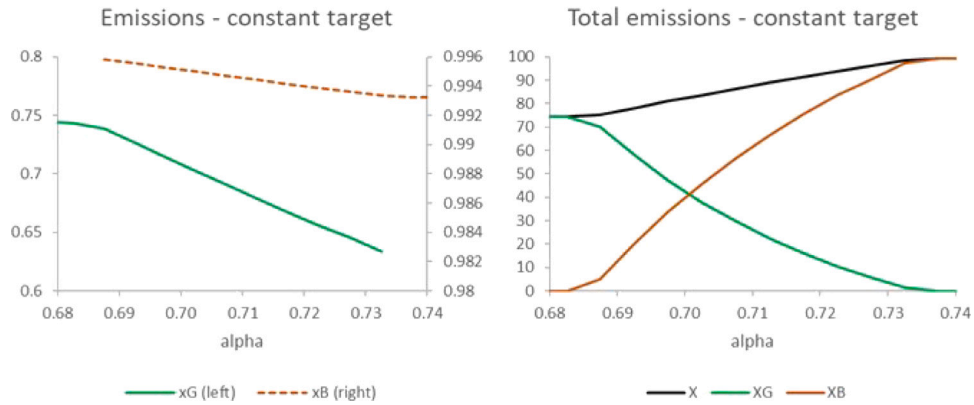

Fig. 6. Left panel: Emissions of green (left axis) and brown (right axis) countries. Right panel: total emissions and total emissions in each group. Values are reported as a function of the ambition parameter α for $r_0 = 0.594$ and $\gamma = 0.00065$. Other parameter values are listed in Table 1.

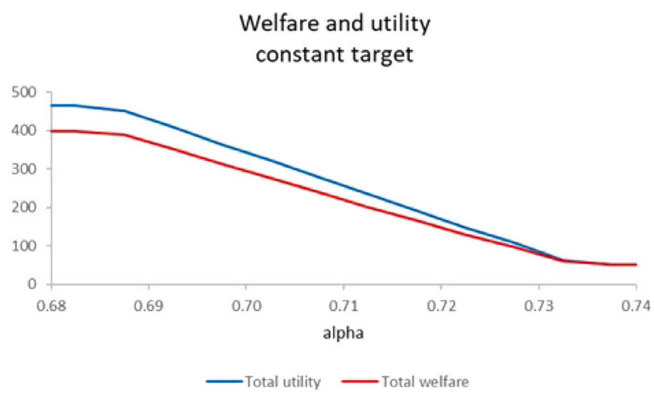

Fig. 7. Total discounted welfare and utility at the steady state as a function of the ambition parameter α for $r_0 = 0.594$ and $\gamma = 0.00065$. Other parameter values are listed in Table 1.

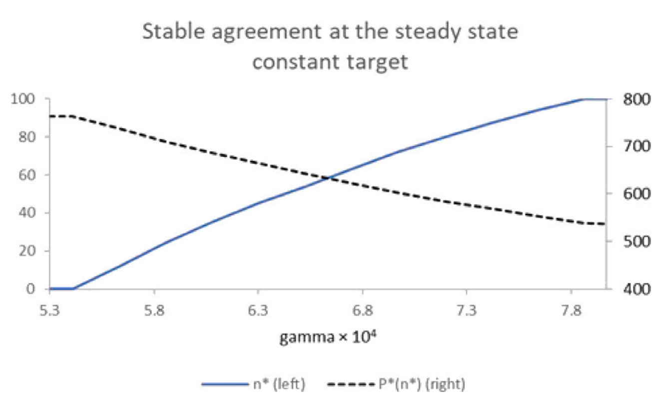

Fig. 8. Left axis: number of green countries in a stable agreement at the steady state. Right axis: pollution stock at the steady state for a stable agreement. Values are represented as a function of the concern level parameter γ for $r_0 = 0.594$ and $\alpha = 0.7$. Other parameter values are listed in Table 1.

The impact of γ on the total emissions is illustrated in the right panel of Fig. 9, showing that participation in the IEA drives the result. While brown countries' individual emissions are increasing with γ, their total contribution is decreasing because the number of brown countries is decreasing with γ. The reverse is true for green countries: as γ increases, the number of green countries increases, and their total contribution to the pollution stock increases. However, since green countries' emissions are significantly lower than those of brown countries, the total of the emissions from both groups decreases with γ.

Finally, Fig. 10 shows that both the total discounted welfare and the total discounted utility are increasing with the concern level parameter γ.

From this analysis we can summarize that a greater environmental awareness improves countries' participation in the IEA, reducing in this way the pollution stock and allowing all countries to achieve a higher level of welfare.

4.2. State-dependent target

In the second set of experiments, the target depends on the current pollution stock. Choosing a state-dependent target is equivalent to identifying a reference strategy, here, linear in the pollution stock ($r_1 > 0$). In the example that follows, the reference strategy is computed by taking the average of the no-agreement and the first-best strategies, so that the reference emissions level is halfway between the two corresponding levels for all P.

Fig. 11 shows that stable agreements can be achieved when the target is decreasing with the pollution stock; for the set of parameter values used in Fig. 11, the solution of the membership game is $n^* = 39$ and the correspondent steady-state pollution stock sits at $P^*(39) = 646$.

Fig. 12 plots the equilibrium value functions (left panel) and emissions (right panel) as a function of the current pollution stock, in the first-best solution, the no-agreement solution ($n = 0$), and the equilibrium solution between green and brown countries as well as their weighted average welfare, at $n = 39$.

Note that, since the equilibrium strategies are linear in P, it is easier for green countries to track a state-dependent than a constant target, so that the self-image of green countries is positive over a large range of possible values of the pollution stock, and the average emissions at equilibrium are close to the target at all P.

In general, the behavior of the equilibrium solution of the emissions game with a state-dependent target is qualitatively similar to the results obtained with a constant one: the green countries' emission strategy is decreasing in the pollution stock, at a higher rate than in the first-best solution, and emissions of green countries are lower, and those of brown countries are slightly higher than in the no-agreement case, except when the pollution stock is very close to 0. As in the constant-target case, the average welfare of countries, not including the self-image portion, is higher than in the no-agreement case for all levels of the pollution stock.

Appendix 6.6 presents the results of sensitivity analyses performed with respect to the ambition of the pledge α and the level of concern γ for the set of parameters used in Fig. 11. Our numerical investigations with various sets of parameters show that the main difference with respect to the constant-target case is that players' strategies and utilities are significantly more sensitive to variations in the self-image parameter values.

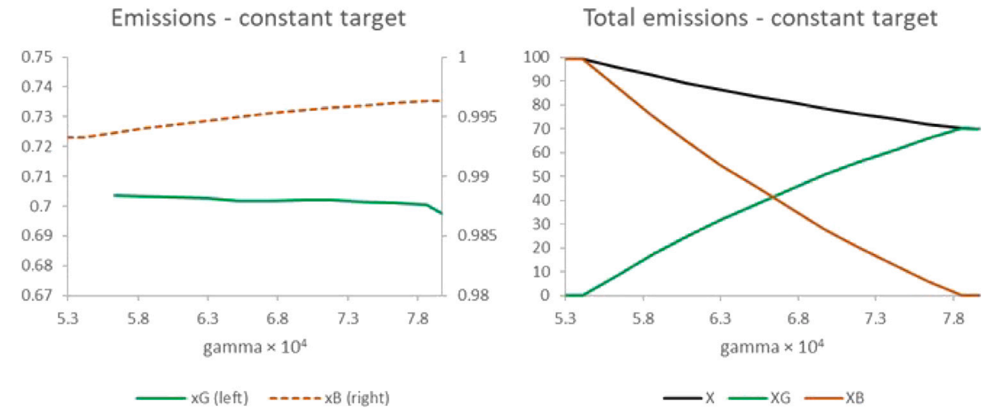

Fig. 9. Left panel: Emissions of green (left axis) and brown (right axis) countries. Right panel: Total emissions and total emissions in each group. Values are represented as a function of the concern level parameter γ for $r_0 = 0.594$ and $\alpha = 0.7$. Other parameter values are listed in Table 1.

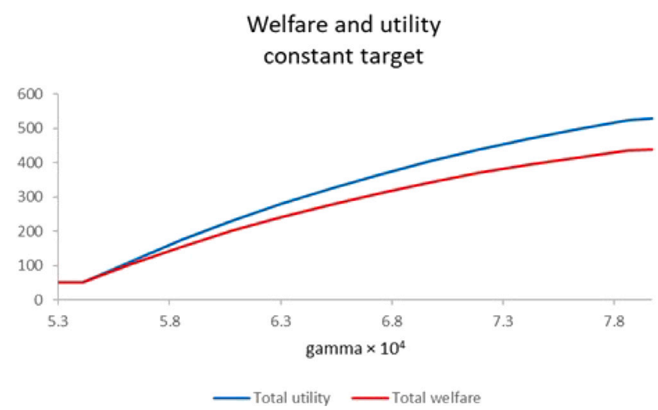

Fig. 10. Total discounted welfare and utility at the steady state, as a function of the concern level parameter γ for $r_0 = 0.594$ and $\alpha = 0.7$. Other parameter values are listed in Table 1.

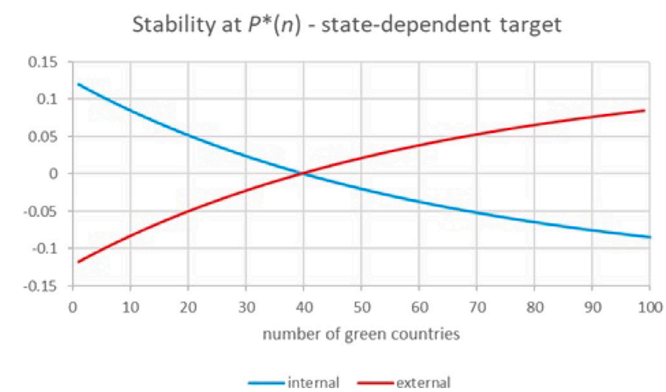

Fig. 11. Stability conditions at $P^*(n)$ as a function of the number of green (signatory) countries n. The self-image function parameter values are $r_0 = 0.887$, $r_1 = 0.0002$, $\alpha = 0.95$, $\gamma = 0.00065$. Other parameter values are listed in Table 1.

Qualitatively, the impact of unilateral variations in the ambition parameter α is similar to what is reported for the constant-target scenario; however, sensitivity analyses of the individual emissions with respect to the concern parameter γ show that, contrary to what happens when the target is constant, both brown and green countries' emissions are increasing in γ (see Figure 19). This result is due to the predominant impact of the decrease in pollution stock over the increase of γ. Nevertheless, as in the constant-target case, the overall impact of the self-image parameters is driven by the size of a stable agreement: since the emissions level of green countries is much lower than that of brown countries, the total emissions decrease with γ as the number of participants in the agreement increases.

To conclude this section, note that all the results and analysis provided are performed at the steady state, where the number of signatories and the stock of pollution no longer change over time. The impact of the ambition and concern parameters on the size of a stable agreement and on players' emissions can also be obtained for different levels of the current stock of pollution. Numerical investigations show that qualitative results about the impact of α and γ on the stable number of signatories, on individual emission levels, and on total emissions, welfare and utility, are robust across feasible values of the state variable.[17]

[17] Recall that total welfare and utility values are computed over an infinite horizon, assuming that players are myopic to the evolution of the number of signatories.

4.3. Trajectories over time

Stability conditions change as the stock of pollution evolves; accordingly, in an open-membership setting,[18] starting from arbitrary initial conditions, both the number of signatory countries and the level of pollution stock can change from period to period. This section provides some examples of such trajectories over time.

At each decision date t, countries observe the current level of the pollution stock and independently decide on their emissions level, according to their membership status. The pollution stock at the next decision date is governed by Eq. (1). As for the number of signatories at the next decision date, we consider three different possibilities.

S1 In Scenario S1, the number of signatory countries is such that the agreement is stable at all decision dates, that is,

$$n_t \in n^*(P_t)$$

where $n^*(P_t)$ represent the set of solutions to the stability conditions (17)–(18) at P_t.

This scenario corresponds to the assumption made in Rubio and Ulph (2007), where the membership game is played and the set of signatory countries is reshuffled at each decision date. In other words, at a given decision date, countries observe

[18] An open membership setting is also used in Rubio and Ulph (2007), de Zeeuw (2008), Breton et al. (2010), Nkuiya (2012), Nkuiya et al. (2015) and Villani and Biancardi (2019), just to mention a few.

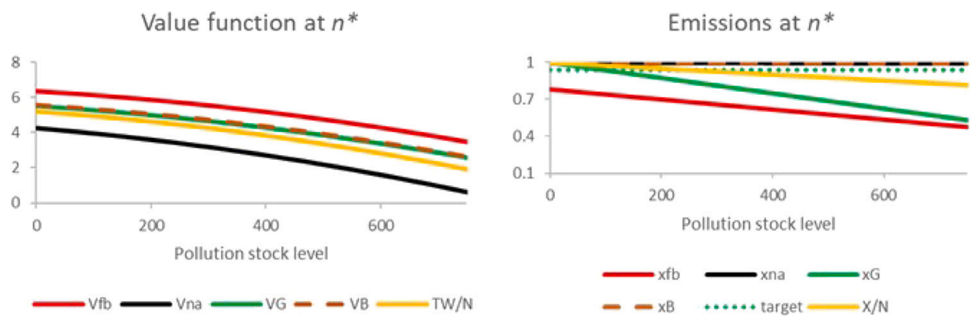

Fig. 12. Equilibrium value function (left panel) and emissions (right panel) of green and brown countries as a function of the pollution stock for $n = 39$, compared to the no-agreement and the first-best solutions. Self-image parameter values are $r_0 = 0.887$, $r_1 = 0.0002$, $\alpha = 0.95$, $\gamma = 0.00065$. Other parameter values are from Table 1.

the level of the pollution stock and make their membership decisions, so that the resulting number of signatory countries solves conditions (17) and (18). The participation level in the previous period does not have any impact on that of the next period.

S2 In Scenario S2, the number of signatory countries updates following a very simple rule: a single country changes its membership status at a decision date when doing so results in a change (increase/decrease) in its total utility:

$$n_{t+1} = \begin{cases} n_t + 1 \text{ if } V^G\left(P_t; n_{t+1}\right) > V^B\left(P_t; n_t\right) \\ n_t - 1 \text{ if } V^B\left(P_t; n_t - 1\right) > V^G\left(P_t; n_t\right) \\ n_t \text{ otherwise.} \end{cases}$$

In this scenario, the membership game is not played, but countries gradually reach a stable agreement through trial and error. Note that in this case, the participation level in the previous period does have an impact on that of the next period.

S3-S4 In Scenarios S3 and S4, the number of signatory countries evolves according to replicator dynamics, where the fitness in each group of countries is measured by comparing the value of their total discounted utility, at the current pollution stock, with an average value in the population. More precisely, the number of signatories in Scenarios S3 and S4 updates respectively using

$$n_{t+1} - n_t = n_t \left(V^G\left(P_t; n_t\right) - \overline{V} \right) \quad \text{(S3)}$$

$$n_{t+1} = n_t \left(\frac{V^G\left(P_t; n_t\right)}{\overline{V}} \right) \quad \text{(S4)}$$

where

$$\overline{V} = n_t V^G\left(P_t; n_t\right) + m_t V^B\left(P_t; n_t - 1\right).$$

This scenario corresponds to the assumption made in Breton et al. (2010), where, as in S2, the membership game is not played, but evolutionary pressures make countries reach a stable IEA at the steady state, and where the speed of change in the number of signatory countries depends on the relative fitness of the two populations. As in Breton et al. (2010), the comparison value $\overline{V}$ is based on the stability conditions (17)–(18), comparing the utility of a green country at the current participation level to what a brown country would achieve if a green country were to defect.

Fig. 13 shows examples of trajectories under the four scenarios for the parameter values used in the example in Fig. 1, where the steady state is 54 signatory countries and a pollution stock of 644. Fig. 13 reports the evolution of the pollution stock and of the number of signatories for 50 time steps, where the four trajectories start from the same initial conditions: in the left panel, both the initial pollution stock and the initial number of countries are lower than the steady state solution and, in the right panel, they are both higher. As expected, all four scenarios produce trajectories that converge to the steady state (trajectories S3 and S4 have not reached the steady state after 50 time steps). The main difference between these trajectories is that the ones generated under Scenario 1 take a direct path towards the steady state, while the trajectories generated by the dynamics under Scenarios 2, 3 and 4 tend to overshoot (*resp. undershoot*) both the stable size and the pollution stock in the left (*resp. right*) panel.

However, there is no reason in Scenarios S2 to S4 to suppose that the initial number of signatories would be the solution of the membership game at $t = 0$; how a set of countries initially decide to participate in negotiations and adhere to an agreement is open to discussion and not treated here. Fig. 14 illustrates the trajectories under the four scenarios starting from a pollution stock of 700 and various initial values for the number of signatories (note that the steady state pollution stock in the noncooperative case is 764 for this set of parameter values).

Fig. 14 also reports the size of a stable agreement as a function of the pollution stock, $n^*(P)$, (dotted line) and the steady-state pollution stock as a function of the number of signatories, $P^*(n)$, (dashed line). The steady-state of the game is the intersection of these two functions. Clearly, the path under S1 coincides with the function $n^*(P)$.[19] From a given point (P, n) situated to the right (resp. left) of the dashed line, the pollution stock will tend to decrease (resp. increase) according to Eq. (1). In the same way, from a given point (P, n) situated over (resp. under) the dotted line, the number of signatories will tend to decrease (resp. increase) according to the stability conditions (17)–(18). The difference between the four trajectories is related to the relative speed of changes in the number of signatories. In Scenario S1, n adjusts instantly to changes in the pollution stock level. In Scenario S2, the adjustment is by at most one country. In Scenarios S3 and S4, the adjustment speed depends on the difference in utility between green and brown countries.[20]

5. Conclusion

It is well recognized in the literature on IEA that stable agreements aimed at reducing global polluting emissions are difficult to forge, mainly because of the free-riding experienced by non-signatory countries. A stable agreement involving a significant number of countries should involve a mechanism allowing a positive outcome for signatories – or a negative one for non-signatories – that enhances the benefit of adhering to the agreement.

In this paper we propose a model contributing to the discussion about the stability of IEA. The main features of this model are:

[19] For a given set of parameters, numerical investigations show that the solution $n^*(P)$ of the membership game is nearly quadratic, convex, and is increasing in P over the range for the pollution stock where $n^*(P) \in (0, N)$.

[20] Note that in the simulation presented in Figs. 13 and 14, n is not required to take integer values in Scenarios S3 and S4.

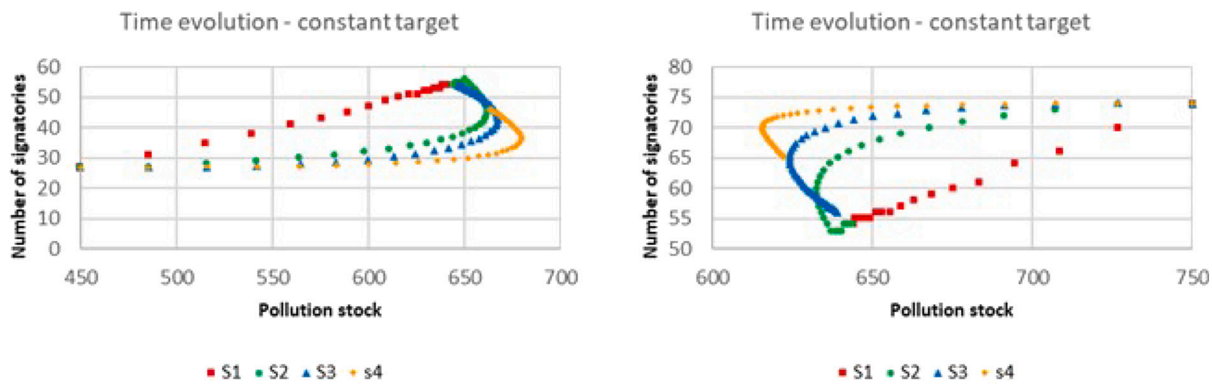

Fig. 13. Trajectories for the pollution stock and the number of signatories over 50 time periods under Scenarios S1, S2, S3 and S4. Initial conditions are $(P_0, n_0) = (450, 27)$ in the left panel and $(750, 74)$ in the right panel. The self-image function parameter values are $r_0 = 0.594$, $\alpha = 0.7$, $\gamma = 6.5 \times 10^{-4}$. Other parameter values are listed in Table 1.

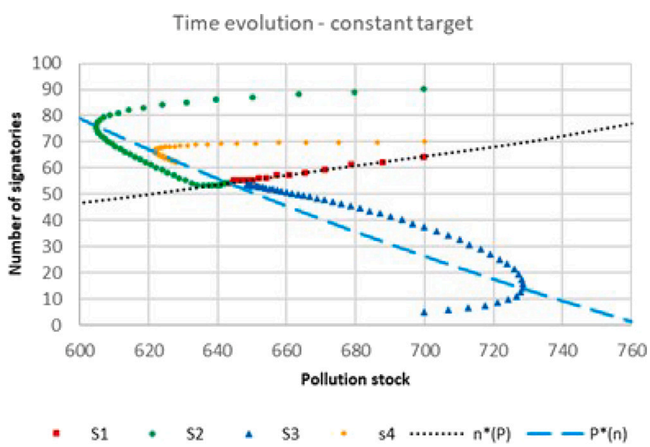

Fig. 14. Trajectories for the pollution stock and the number of signatories over 50 time periods under Scenarios S1, S2, S3 and S4. Initial pollution stock is $P_0 = 700$. Initial number of signatories is, respectively, $n_0 = 64$, 90, 5, and 70 for S1, S2, S3 and S4. The self-image function parameter values are $r_0 = 0.594$, $\alpha = 0.7$, $\gamma = 6.5 \times 10^{-4}$. Other parameter values are listed in Table 1.

- the agreement is about committing to attain a common target on individual emissions;
- all countries choose their emissions strategies independently;
- countries that sign the agreement value their self-image arising from the commitment taken. The self-image component in their utility can be either negative, arising from shame due to not complying with their commitment, or positive, due to pride in contributing to the environmental cause.

Assuming that an agreement has been reached on the common target, our model investigates the stability of the agreement and the equilibrium emissions strategies of countries, in a dynamic setting where emissions contribute to the global pollution stock. Our investigation shows that, when countries value their self-image, stable IEAs with meaningful participation are achievable, and are effective in terms of pollution stock and global welfare.

To better understand the role of self-image in the stability and efficiency of agreements, we analyze how these results are affected by two fundamental self-image parameters, namely, the ambition of the pledge, and the level of concern about environmental issues in signatory countries. We find that

- there is a trade-off between ambition and participation, which results in a negative repercussion of the ambition level on environmental and economic outcomes;
- there is a fortifying relationship between participation and countries' awareness of the risks and dangers of climate change, resulting in a positive impact of countries' concern on environmental and economic outcomes.

These results are robust to the form of the target (constant or indexed on the current pollution stock level) and to the current level of pollution.

The main implication of our investigations can be summarized by concluding that broad participation of countries in an IEA is the key driver to reducing the stock of pollution and increasing the overall welfare. This reinforces the notion that a collective action is needed to tackle global problems, such as climate change. We submit that a large collective action, induced by a modest level of ambition, is more effective than a small number of countries committing to highly curbing their emissions.

On the other hand, we also find that at some point where full participation is achieved, further reducing the ambition of pledges or increasing the weight of the self-image in countries' utility could have a negative impact on overall welfare and environmental condition.

As a future research direction, it remains to investigate the choice of the target, or, equivalently, the ambition of the pledge, which is a crucial design parameter in the conception of an IEA based on a common commitment to attain some emission level.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ecolecon.2023.107869.

References

Andreoni, J., 1990. Impure altruism and donations to public goods: A theory of warm-glow giving. Econ. J. 100 (401), 464–477.

Bahn, O., Breton, M., Sbragia, L., Zaccour, G., 2009. Stability of international environmental agreements: An illustration with asymmetrical countries. Int. Trans. Oper. Res. 16 (3), 307–324.

Barrett, S., 1994. Self-enforcing international environmental agreements. Oxf. Econ. Pap. 46, 878–894.

Barrett, S., 1997. The strategy of trade sanctions in international environmental agreements. Resour. Energy Econ. 19 (4), 345–361.
Barrett, S., 2003. Environment and Statecraft: The Strategy of Environmental Treaty-Making. Oxford University Press, Oxford.
Barrett, S., Dannenberg, A., 2016. An experimental investigation into 'pledge and review' in climate negotiations. Clim. Change 138 (1), 339–351.
Benchekroun, H., Van Long, N., 2016. Status concern and the exploitation of common pool renewable resources. Ecol. Econom. 125, 70–82.
Botteon, M., Carraro, C., 1997. Burden-sharing and coalition stability in environmental negotiations with asymmetric countries. In: Carraro, C. (Ed.), International Environmental Negotiations: Strategic Policy Issues. Edward Elgar, Cheltenham, pp. 26–55.
Brekke, K.A., Kverndokk, S., Nyborg, K., 2003. An economic model of moral motivation. J. Public Econ. 87 (9–10), 1967–1983.
Breton, M., Garrab, S., 2014. Evolutionary farsightedness in international environmental agreements. Int. Trans. Oper. Res. 21 (1), 21–39.
Breton, M., Sbragia, L., Zaccour, G., 2010. A dynamic model for international environmental agreements. Environ. Resour. Econ. 45 (1), 25–48.
Card, D., Mas, A., Moretti, E., Saez, E., 2012. Inequality at work: The effect of peer salaries on job satisfaction. Amer. Econ. Rev. 102 (6), 2981–3003.
Carraro, C., Marchiori, C., Oreffice, S., 2009. Endogenous minimum participation in international environmental treaties. Environ. Resour. Econ. 42, 411–425.
Carraro, C., Siniscalco, D., 1993. Strategies for international protection of the environment. J. Public Econ. 52, 309–328.
Carraro, C., Siniscalco, D., 1997. R & D cooperation and the stability of international environmental agreements. In: Carraro, C. (Ed.), International Environmental Negotiations: Strategic Policy Issues. Edward Elgar, Cheltenham, pp. 71–96.
Colombo, L., Labrecciosa, P., 2018. Consumer surplus-enhancing cooperation in a natural resource oligopoly. J. Environ. Econ. Manag. 92, 185–193.
d'Aspremont, C., Jacquemin, A., Gabszewicz, J.J., Weymark, J.A., 1983. On the stability of collusive price leadership. Canad. J. Econ. 1, 7–25.
de Zeeuw, A., 2008. Dynamics effects on the stability of international environmental agreements. J. Environ. Econ. Manag. 55, 163–174.
Eyckmans, J., Kverndokk, S., 2010. Moral concerns on tradable pollution permits in international environmental agreements. Ecol. Econom. 69, 1814–1823.
Falkner, R., 2016. The Paris agreement and the new logic of international climate politics. Int. Aff. 92 (5), 1107–1125.
Fehr, E., Schmidt, K.M., 1999. A theory of fairness, competition, and cooperation. Q. J. Econ. 114 (3), 817–868.
Finus, M., Rübbelke, D.T.G, 2013. Public good provision and ancillary benefits: The case of climate agreements. Environ. Resour. Econ. 56, 211–226.
Frank, R.H., 2005. Positional externalities cause large and preventable welfare losses. Am. Econ. Rev. 95 (2), 137–141.
Grüning, C., Peters, W., 2010. Can justice and fairness enlarge international environmental agreements? Games 1 (2), 137–158.
Hartmann, P., Eisend, M., Apaolaza, V., D'Souza, C., 2017. Warm glow vs. altruistic values: how important is intrinsic emotional reward in proenvironmental behavior? J. Environ. Psychol. 52, 43–55.
Hoel, M., 1992. International environment conventions: The case of uniform reductions of emissions. Environ. Resour. Econ. 2 (2), 141–159.
Hoel, M., Schneider, K., 1997. Incentives to participate in an international environmental agreement. Environ. Resour. Econ. 9, 153–170.
Khaneman, D., Knetsch, J.L., 1992. Valuing public goods: the purchase of moral satisfaction. J. Environ. Econ. Manag. 22, 57–70.
Lange, A., 2006. The impact of equity preferences on the stability of international environmental agreements. Environ. Resour. Econ. 34, 247–267.
Lange, A., Vogt, C., 2003. Cooperation in international environmental negotiations due to a preference for equity. J. Public Econ. 87, 2049–2067.
Marchiori, C., Dietz, S., Tavoni, A., 2017. Domestic politics and the formation of international environmental agreements. J. Environ. Econ. Manag. 81, 115–131.
Nkuiya, B., 2012. The effects of the length of the period of commitment on the size of stable international environmental agreements. Dyn. Games Appl. 2, 411–430.
Nkuiya, B., Marrouch, W., Bahel, E., 2015. International environmental agreements under endogenous uncertainty. J. Public Econ. Theory 17 (5), 752–772.
Petrakis, E., Xepapadeas, A., 1996. Environmental consciousness and moral hazard in international agreements to protect the environment. J. Public Econ. 60, 95–110.
Rubio, S.J., Ulph, A., 2007. An infinite-horizon model of dynamic membership of international environmental agreements. J. Environ. Econ. Manag. 54 (3), 296–310.
Tingley, D., Tomz, M., 2021. The Effects of Naming and Shaming on Public Support for Compliance with International Agreements: An Experimental Analysis of the Paris Agreement. International Organization, page 1 of 24.
van der Pol, H.P., van Ierland, E., 2012. Can altruism stabilise international climate agreements? Ecol. Econ. 81, 112–120.
Van Long, N., 2020. Warm glow and the transmission of pro-socialness across generations. J. Public Econ. Theory 22 (2), 371–387.
Van Long, N., 2021. Dynamic games of common-property resource exploitation when self-image matters. pp. 81–107.
Venhoeven, L.A., Bolderdijk, J.W., Steg, L., 2016. Why acting environmentally-friendly feels good: Exploring the role of self-image. Front. Psychol. 7, 1846.
Villani, G., Biancardi, M., 2019. An evolutionary game approach in international environmental agreements with R & D investments. Comput. Econ. 54, 1027–1042.
Weikard, H.P., Wangler, L., Freytag, A., 2015. Minimum participation rules with heterogeneous countries. Environ. Resour. Econ. 62 (4), 711–727.
Wirl, F., 2011. Global warming with green and brown consumers. Scand. J. Econ. 113 (4), 866–884.