**Les Cahiers du GERAD** ISSN: 0711–2440

# Self-image and the stability of international environmental agreements

M. Breton, L. Sbragia

G–2021–47

August 2021



**Citation suggérée :** M. Breton, L. Sbragia (Août 2021). Self-image and the stability of international environmental agreements, Rapport technique, Les Cahiers du GERAD G– 2021–47, GERAD, HEC Montréal, Canada.

**Avant de citer ce rapport technique,** veuillez visiter notre site Web (https://www.gerad.ca/fr/papers/G-2021-47) afin de mettre à jour vos données de référence, s'il a été publié dans une revue scientifique.

**Suggested citation:** M. Breton, L. Sbragia (August 2021). Self-image and the stability of international environmental agreements, Technical report, Les Cahiers du GERAD G–2021–47, GERAD, HEC Montréal, Canada.

**Before citing this technical report,** please visit our website (https://www.gerad.ca/en/papers/G-2021-47) to update your reference data, if it has been published in a scientific journal.

La publication de ces rapports de recherche est rendue possible grâce au soutien de HEC Montréal, Polytechnique Montréal, Université McGill, Université du Québec à Montréal, ainsi que du Fonds de recherche du Québec – Nature et technologies.

The publication of these research reports is made possible thanks to the support of HEC Montréal, Polytechnique Montréal, McGill University, Université du Québec à Montréal, as well as the Fonds de recherche du Québec – Nature et technologies.

Dépôt légal – Bibliothèque et Archives nationales du Québec, 2021
– Bibliothèque et Archives Canada, 2021

Legal deposit – Bibliothèque et Archives nationales du Québec, 2021
– Library and Archives Canada, 2021

**GERAD** HEC Montréal
3000, chemin de la Côte-Sainte-Catherine
Montréal (Québec) Canada H3T 2A7

**Tél. : 514 340-6053**
Téléc. : 514 340-5665
info@gerad.ca
www.gerad.ca

# Self-image and the stability of international environmental agreements

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

[a] *GERAD & Department of Decision Sciences, HEC Montréal, Montréal (Qc), Canada, H3T 1J4*

[b] *Durham University Business School, Durham DH1 3LB, United Kingdom*

michele.breton@hec.ca

**August 2021**
**Les Cahiers du GERAD**
**G–2021–47**

**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.

**Keywords:** Dynamic game, climate change, stability

**Résumé :** Nous étudions la stabilité des accords environnementaux internationaux portant sur une cible commune pour le niveau d'émissions polluantes. En signant l'accord, les parties développent un sens des responsabilités vis-à-vis de l'engagement pris, acquérant une image de soi qui contribue à leur utilité.

Nous analysons un jeu dynamique en deux étapes où tous les pays agissent de manière indépendante. Nous étudions comment deux composantes fondamentales du modèle, à savoir le niveau (ambition) de l'engagement et l'importance relative accordée par les signataires au respect de leur engagement, affectent la stabilité et l'efficacité de l'accord en termes de bien-être global et d'émissions totales.

Nous constatons que la participation à l'accord constitue le moteur clé des résultats. La participation diminue avec l'ambition de l'engagement et augmente avec le niveau de préoccupation des pays pour les questions environnementales.

**Mots clés :** Jeux dynamiques, changement climatique, stabilité

**Acknowledgements:** Research supported by NSERC (Canada).

# 1 Introduction

During the last two years, several individual nation states, countries, territories, and councils published declarations on the climate emergency (for instance, the UK in May 2019, Canada and France in June 2019, the EU in November 2019, Japan in November 2020), indicating that climate change is an important issue in the political agenda of countries worldwide. The year 2021 is of particular importance in the fight against climate change, as the parties to the Paris Agreement will meet in November in Glasgow to review their commitments to reduce greenhouse gas (GHG) emissions in order to attain the objective stated in Article 2 of the Agreement.[1]

In the long history of international cooperation to tackle the problem of climate change, beginning in 1988 with the establishment of the Intergovernmental Panel on climate change (IPCC), the most recent milestone can be considered to be the Paris Agreement (2015). Among its many distinguishing features, the Paris Agreement (a) retains the "principle of equity and common but differentiated responsibilities and respective capabilities" in taking actions against climate change (Article 2); (b) fixes a target in terms of temperature increase (Article 2), giving parties the flexibility to decide on the way to reach it (bottom-up approach); (c) requires parties to declare their nationally determined contributions (NDCs) to the global response to climate change (Article 3) (Note that these contributions are not binding but must reflect the parties' "highest possible ambition" (Article 4)); (d) calls for commitments to be updated every five years, showing a progression beyond the parties' then-current NDC (Article 4); (e) demands a high level of transparency (Articles 4 and 13), which can be considered a form of control or peer pressure; and (f) does not include any form of punishment; however it incorporates the so-called *Global Stocktake* (Article 14), which assesses the implementation of the Agreement and the collective progress towards achieving its purpose.

The outcome of the Paris Agreement left experts with mixed feelings: on the one hand, it achieved broad participation by countries (195 parties signed, and 189 ratified the Agreement);[2] on the other hand, the pledges made by countries at Paris to reduce emissions seemed insufficient to meet the agreed-upon temperature target.[3]

In November, parties to the Agreement will find themselves once again at a crossroads, where they will have to announce their new pledges for the next period, putting their "ambition" under the spotlight.

In this paper we contribute to the literature on international environmental agreements (IEAs) by proposing a stylized model that builds on three key features of the Paris Agreement: ambition, transparency, and freedom of action. Using this model, we examine how the ambition and commitment levels of signatory countries affect the stability and the efficiency of an agreement in terms of global welfare and total emissions.

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 ("emissions" game). Countries with different participation statuses (signatories vs nonsignatories) show heterogeneous preferences in the emissions game.

We assume that the IEA is about some agreed-upon (common) target, whose value depends on the level of ambition,[4] so that, when a country becomes a signatory, it publicly commits to that target. The transparency requirement works both as a form of control on a country's progress towards its commitment and as a source of peer pressure. In this way, making a public commitment to an

[1] "holding the increase in the global average temperature to well below 2 °C above preindustrial levels and pursuing efforts to limit the temperature increase to 1.5 °C above preindustrial levels" (Article 2).

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

[3] The first Global Stocktake will take place from 2021 to 2023.

[4] The process of how this agreed-upon target is reached is not modeled in this paper.

agreed-upon target develops a sense of responsibility towards that target in the signatory country, or a self-perception.

In particular, we assume that the utility of a signatory country is composed of both its monetary welfare and a self-image related to the commitment taken. The self-image portion is characterized by its weight in the signatory's utility; a reference point; and the ambition of the pledge. The target depends on both the reference point and the ambition of the pledge. If the signatory country is able to overcomply with respect to the target, it experiences a sense of pride (warm glow), while if it is not able to respect its commitment, it feels a sense of shame. We assume that the weight of the self-image depends on the seriousness of the global environmental problem, on a country's awareness and concern about it, and on the ambition of the country's commitment. In terms of the reference point, we consider two possibilities: a simple benchmark (a given emissions level)[5] and, to reflect the possible revisions included in the Paris Agreement, a more sophisticated reference that adapts to the current situation (e.g. to the accumulated stock of pollutants).

We first investigate whether this setting can generate stable agreements of meaningful size and impact. We then move our attention to an analysis of the relationship between the ambition of the pledge and the size of the stable IEA, and its environmental and economic consequences. This issue is related to Article 4 of the Paris Agreement, and is derived from the current debate about the ambition of the new NDCs countries must communicate at COP26.[6] Finally, we investigate how the relative weight of self-image within the countries' utility affects the IEA's stability. This investigation is motivated by the many different initiatives launched on various media to inform and educate people about climate change and its associated risks, on the road to COP26.

Our first contribution is to show that, when countries care about their self-image, 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. 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 also achieved with a greater awareness of the dangers and risks of the global environmental problem. Our findings are robust to the model's parameter values and to the type of reference point.

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 & Siniscalco (1993), and Barrett (1994), this literature has developed abundantly over the years; see for instance, to mention only a few, Petrakis & Xepapadeas (1996), Botteon & Carraro (1997), Hoel & Schneider (1997), Barrett (1997), Carraro & Siniscalco (1997), and Carraro et al. (2009) in a static framework; Rubio & Ulph (2007), De Zeeuw (2008), and Breton et al. (2010) in a dynamic context; and, more recently, Finus & 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 & Vogt (2003) and Lange (2006), countries share a sense of equity; in Grüning & Peters (2010), countries' preferences incorporate justice and fairness; and in van der Pol et al. (2012), altruism affects all countries, albeit in different ways. The 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

[5]For instance, the UK Prime Minister legislated the new target to cut carbon emissions by 78% compared to 1990 levels, by the year 2035.
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
US President Biden promised to halve his country's greenhouse gas emissions from 2005 levels, by the year 2030.

[6]https://www.gov.uk/government/publications/g7-climate-and-environment-ministers-meeting-may-2021-communique/g7-climate-and-environment-ministers-communique-london-21-may-2021

freedom to reach it in the way most convenient to them. This has two consequences: the first is that all countries, including signatory countries, compete against each other in the emissions game; the second is that, since nonsignatory countries do not make any commitment, their utility does not include any self-image concerns.

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 & van Long 2016); models with altruistic players, who include other players' welfare in their objective function (as in Fehr & Schmidt 1999 and Colombo & 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 & 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 Barret (2003), that is, the possible trade-off between the breadth of international 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

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, 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

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

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

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

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

The above assumptions characterize the standard stylized model commonly used in the IEA literature,[7] 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 adhereto the agreement, an 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 Equation (2), and a self-image component, related to the extent to which a country complies with the pledge making up the agreement.

[7] 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.

## 2.1 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 (sense of shame). 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 weight of a positive or of a negative deviation with respect to the target are the same.

The target is defined by

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

where $r$ is the reference point and $\alpha$ is the ambition. The reference point represents a common norm. This can be a constant (e.g. the business-as-usual, the current, the preindustrial, or the first-best emissions level) or it can be a function of the state of the system (here, the pollution level). The ambition parameter $\alpha$ is a positive constant that situates the target with respect to the reference point. Since we are considering a public bad, the target is decreasing in the ambition of the pledge, so that a more ambitious pledge requires countries to attain an emissions level that is lower than a less ambitious one. Assuming that the reference point is linear in the pollution level yields the following general form for the self-image function for country $i$ in time period $t$

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

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

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

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

where $\lambda_t$ is the weight of the self-image. This weight characterizes the relative importance of economic welfare and self-image in the utility of green countries.

It is reasonable to assume that the relative importance of self-image is increasing with the seriousness of the global environmental problem, with countries' awareness and concern with respect to climate change, and with the ambition of their pledge. We use the simple specification

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

to model the impact of these three factors.

According to Equation 3, the weight of self-image is proportional to the stock of accumulated pollution, representing the seriousness of the environmental problem. The proportion coefficient is increasing with $\gamma$, which represents awareness of climate change and of the risks associated with it, which can be enhanced through education campaigns and information dissemination. Finally, at a given level of pollution and awareness, the weight of self-image is also increasing with $\alpha$.[8] With respect to a given reference point, a more ambitious pledge generates more warm glow than a less ambitious one.

[8] Note that $\alpha$ is a design parameter. By making the weight of the self-image dependent on the ambition, we avoid the possibility that agreements will generate a large positive utility by selecting easily attainable targets, such as the status quo.

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

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

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

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 revenue and environmental damage functions no longer symmetric.

# 3 Dynamic game setting

In this section we develop and solve the two-stage game played by countries. 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.

## 3.1 Emissions game

We 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.

### 3.1.1 Reaction function

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

Consider a single country, with an immediate utility that takes the general form

$$U_t\left(x_t, P_t\right) = c_1 x_t^2 + c_2 x_t + c_3 P_t x_t + c_4 P_t^2 + c_5 P_t.$$

The optimization problem faced by this country is then

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

where $O_t$ is 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 total emissions level is linear in $P$, the best response of a single country is also linear in $P$.

**Proposition 1.** If the total of the emissions by all countries is linear in the pollution stock

$$X \equiv \sum_{i=1}^{N} x_i = Q_0 + Q_1 P,$$

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

$$x(P) = q_0 + q_1 P,$$

provided that there exist constants $b_0$, $b_1$, and $b_2$ satisfying

$$b_2 = c_4 - \frac{1}{4c_1} c_3^2 + \beta M_1^2 b_2 \frac{c_1 + \beta b_2}{c_1} \tag{6}$$

$$b_1 = c_5 + \frac{-c_2 c_3 + 2\beta^2 M_1 M_2 b_2}{2c_1} + \beta M_1 M_2 \tag{7}$$

$$b_0 = \frac{M_2^2 \beta^2 - c_2^2}{4c_1} + \beta \left(b_0 + Q_0 M_2\right) \tag{8}$$

$$0 > c_1 + b_2 \beta \tag{9}$$

$$M_1 = Q_1 + \delta \tag{10}$$

$$M_2 = b_1 + 2Q_0 b_2. \tag{11}$$

**Proof.** See Appendix A.1. □

Note that $Q_0$ and $Q_1$ are global values that need to be defined according to the equilibrium solution concept.

### 3.1.2 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 $m + n = N$. Parameters pertaining to green and brown countries are indexed respectively by $G$ and $B$, where

$$\begin{aligned}
c_1^G &= c_1^B = -\frac{1}{2} \\
c_2^G &= c_2^B = 1 \\
c_3^G &= -\alpha\gamma \ ; \ c_3^B = 0 \\
c_4^G &= \gamma r_1 - d_2 \ ; \ c_4^B = -d_2 \\
c_5^G &= \gamma r_0 - d_1 \ ; \ c_5^B = -d_1.
\end{aligned}$$

In Proposition 1, we showed that there exists a Nash equilibrium to the emissions game in linear feedback stationary strategies if we can obtain parameters $b_k^G$ and $b_k^B$, $k \in \{0, 1, 2\}$, characterizing the reaction function of each group of countries, satisfying the system (6)–(11) involving $Q_0$ and $Q_1$. Proposition 2 characterizes the global values $Q_0$ and $Q_1$ corresponding to a Nash equilibrium.

**Proposition 2.** If there exists a linear Nash equilibrium feedback strategy to the emissions game, the global values $Q_0$ and $Q_1$ satisfy

$$\begin{aligned}
Q_0 &= \frac{N + m\beta b_1^B + n\beta b_1^G}{4\Phi\beta + 1} \\
Q_1 &= \frac{-\alpha\gamma n - 4\Phi\beta\delta}{4\Phi\beta + 1}
\end{aligned}$$

where

$$\Phi \equiv -\frac{m b_2^B + n b_2^G}{2}. \tag{12}$$

**Proof.** See Appendix A.2. □

For a fixed $\Phi$, the value of $M_1$ depends only on given parameters:

$$M_1 = \frac{\delta - n\alpha\gamma}{4\Phi\beta + 1}.$$

Accordingly, one can solve for the value function coefficients $b_2^G$ and $b_2^B$ by finding the roots of two independent quadratic equations:

$$-2\left(\beta M_1 b_2^G\right)^2 + b_2^G\left(\beta M_1^2 - 1\right) + \gamma r_1 - d_2 + \frac{1}{2}\alpha^2\gamma^2 = 0 \tag{13}$$

$$-2\left(\beta M_1 b_2^B\right)^2 + b_2^B\left(\beta M_1^2 - 1\right) - d_2 = 0. \tag{14}$$

A numerical solution to the conditions (6) for green and brown players can be sought using fixed-point iteration, using Equation (12) to update the value of $\Phi$ (note that multiple solutions may exist).[9]

If there exists a pair of values for parameters $b_2^G$ and $b_2^B$ satisfying the system (12)–(14), then it is straightforward to obtain the corresponding equilibrium solution of the emissions game by solving a system of linear equations (see Appendix A.3). The equilibrium solution is given by

$$\begin{aligned} x^B &= 1 + \beta\left(b_1^B + 2Q_0 b_2^B\right) + 2P\beta M_1 b_2^B \\ x^G &= 1 + \beta\left(b_1^G + 2Q_0 b_2^G\right) + P\left(-\alpha\gamma + 2\beta M_1 b_2^G\right) \end{aligned}$$

where

$$\begin{aligned} b_1^B &= \frac{A^B n\beta b_2^B\left(\gamma\left(\alpha - r_0\right) + d_1 + A^G N b_2^G\right) + \frac{1}{2}\left(d_1 + A^B N b_2^B\right)\left(A^G\left(2m\beta b_2^B - 1\right) - 2\right)}{\left(A^G K^B - 1\right)\left(A^B K^G - 1\right) - A^B A^G m n\beta^2 b_2^B b_2^G} \\ b_1^G &= \frac{A^G m\beta b_2^G\left(d_1 + A^B N b_2^B\right) + \frac{1}{2}\left(\gamma\left(\alpha - r_0\right) + d_1 + A^G N b_2^G\right)\left(A^B\left(2n\beta b_2^G - 1\right) - 2\right)}{\left(A^G K^B - 1\right)\left(A^B K^G - 1\right) - A^B A^G m n\beta^2 b_2^B b_2^G} \end{aligned}$$

$$\begin{aligned} A^B &= 2\beta M_1 \frac{2\beta b_2^B - 1}{4\Phi\beta + 1}; A^G = 2\beta M_1 \frac{2\beta b_2^G - 1}{4\Phi\beta + 1} \\ K^B &= m\beta b_2^B - \frac{1}{2}; K^G = n\beta b_2^G - \frac{1}{2}. \end{aligned}$$

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 noncooperative 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 $na$.

For a given $n$, the steady state $P^*(n)$ of the pollution level corresponding to the equilibrium solution of the emissions game is defined by

$$\delta P_t + Q_0 + Q_1 P_t = P_t,$$

yielding

$$P^*(n) = \frac{Q_0}{1 - M_1}. \tag{15}$$

[9]Note that the value functions are expected to be concave decreasing for realistic parameter values, so a negative root should be selected.

### 3.1.3 First-best solution

In the absence of an agreement, the cooperative equilibrium, or first-best solution, corresponds to the emissions strategy that maximizes the total welfare of all countries, that is,

$$\begin{aligned} \max_x & \left\{ \sum_{t=0}^{\infty} \beta^t U_t^B \left(x_t, P_t\right) \right\} \\ \text{s.t.} & \\ P_t =& \delta P_{t-1} + N x_t. \end{aligned}$$

The first-best is the solution of a dynamic optimization problem, and it is straightforward to show that the value function is quadratic and that the first-best emissions strategy is linear in the current pollution stock $P$ (see Appendix A.4). The first-best equilibrium is given by

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

where

$$\begin{aligned} b_2^{fb} =& \frac{1 - \beta\left(\delta^2 + 2N^2 d_2\right) - \sqrt{\left(\beta\delta^2 + 2N^2\beta d_2 - 1\right)^2 + 8N^2\beta d_2}}{4N^2\beta} \\ b_1^{fb} =& \frac{2N\beta b_2^{fb}\left(\delta + N d_1\right) - d_1}{1 - \beta\left(\delta + 2N^2 b_2^{fb}\right)}. \end{aligned}$$

## 3.2 Membership game

To solve the membership game,[10] 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. In a dynamic framework, players in a given group compare their total discounted utility 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.

Now, the value of the emissions game is a function of the current level of the pollution stock $P$. At a given $P$, players compare their utility in or out of the agreement; therefore, the number of signatories is stable at $P$ when

$$V^G(P;n) \geq V^B(P;n-1) \tag{16}$$

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

Clearly, the size of a stable agreement depends on the current pollution level.

Note that our setting can accommodate the possibility of both closed- and open-membership situations. In the closed-membership case, the membership game is played once, at $t = 0$, and the number of participating countries remains constant afterwards. The size of the stable agreement is then the solution of Equations (16)–(17) at $P_0$. In the open-membership case, countries can decide to adhere or leave the agreement at any time, so that the number of participating countries can change over time. Note that in the open-membership case, we assume that players are myopic with respect to the time evolution of the proportion of green countries.[11]

[10] 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.

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

In this paper, we investigate the stability of agreements at the steady-state pollution level, that is, we look for a pair $(n^*, P^*)$ such that Equations (16)–(17) are satisfied at

$$P^*(n^*) = \frac{Q_0(n^*)}{1 - M_1(n^*)}.$$

In an open-membership setting, the pair $(n^*, P^*)$ is a steady state, in terms of the pollution stock and of the number of signatory countries. In a closed-membership setting, when $P_0 = P^*(n^*)$, the solution of the membership game is such that the pollution stock is stable over time, and therefore, there is no incentive for the players to reopen membership negotiations. In all other cases, the stability conditions will not be satisfied over time.

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

In the solution of the emissions game, emissions by countries are not restricted to being positive. Depending on the parameter values defining the welfare functions ($\delta$, $\beta$, $d_1$, $d_2$) and the self-image functions ($r_0$, $r_1$, $\gamma$, $\alpha$), and on the number of countries ($N$, $n$), it can happen that the best response in one (or both) groups of players is negative. While negative emissions could theoretically be possible, in our numerical investigation we rule out the existence of actions that could decrease the pollution stock (apart from natural decay). This is related to our assumption that emissions are proportional to the production level, and therefore that a player's revenue would vanish when its emissions are equal to 0.

We use the same base-case parameter values for $N$, $\delta$, $\beta$, $d_1$, 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 under this restriction.

**Table 1: Base-case parameter values.**

| $N$ | $\delta$ | $\beta$ | $d_1$ | $d_2$ |
|---|---|---|---|---|
| 100 | 0.87 | 0.95 | $1.0 \times 10^{-5}$ | $3.0 \times 10^{-7}$ |

## 4.1 Constant reference point

In the first set of experiments, the reference point is a given emissions level $r_0$, so that the self-image function is defined by

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

Figure 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 Equation (15). Figure 1 plots the value of

i) $V^G(P^*(n); n) - V^B(P^*(n); n)$ (difference),
ii) $V^G(P^*(n); n) - V^B(P^*(n); n-1)$ (internal stability),
iii) $V^B(P^*(n); n) - V^G(P^*(n); n+1)$ (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, the utilities of the green and brown countries are equal at $P^*(n)$ when $n = 54$. At $n = 54$, both internal and external stability conditions are met, 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.

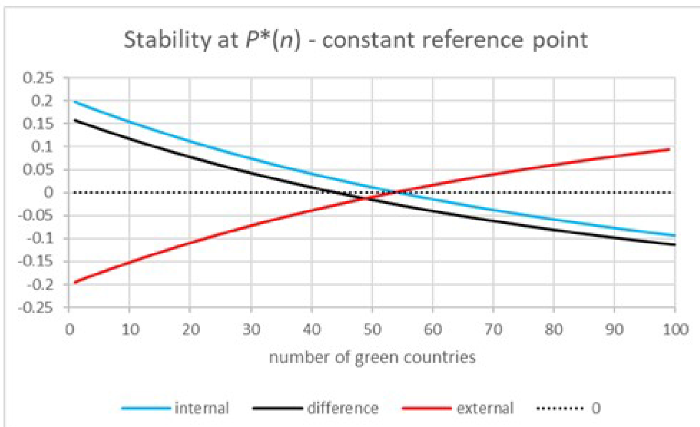


**Figure 1: Stability conditions at $P^*(n)$ as a function of the number of green (signatory) countries $n$.** The self-image function parameter values are $r_0 = 0.594$, $\alpha = 0.7$, $\gamma = 6.5 \times 10^{-4}$. Other parameter values are listed in Table 1.

Figure 2 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 at $n = 54$.

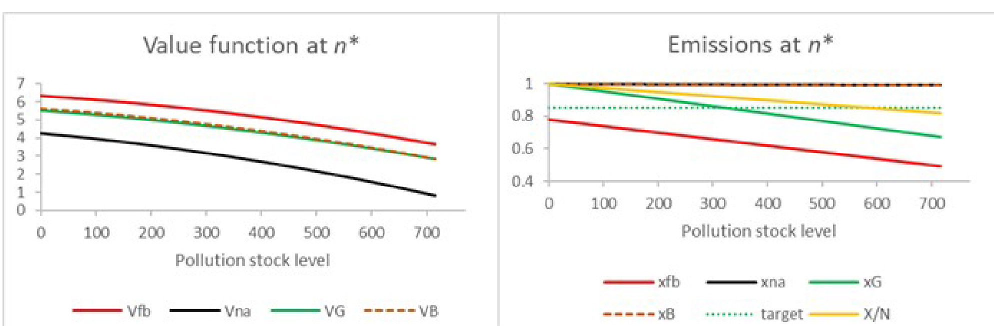


**Figure 2: Equilibrium value function (left panel) and emissions (right panel) of green and brown countries as a function of the pollution stock for $n = 54$, compared to the first-best and the no-agreement solutions.** Self-image parameter values are $r_0 = 0.594$, $\alpha = 0.7$, and $\gamma = 6.5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

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. Emissions of green countries are generally lower, and those of brown countries are slightly higher, than in the no-agreement case. 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.

Note that because the emissions of brown countries are higher than in the no-agreement solution, and because the emissions of green and brown countries coincide at $P = 0$, it can happen, for very low values of $P$, that global emissions at $n^*$ are higher than the no-agreement total emissions level (see Figure A.1 in Appendix A.5.1). This is not the case at equilibrium, $P^*(n^*) = 644$, where global emissions are well below the no-agreement level.

Furthermore, it is interesting to note that parameters can be adjusted such that the outcome of a partial agreement can even outweigh the first-best solution from both the economic and environmental points of view. In particular, this is possible when both the ambition ($\alpha$) and the concern ($\gamma$) parameters take large values, as illustrated in Appendix A.5.2.[12]

We can thus 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, in some instances, with respect to the first-best 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 ($\alpha$) and the concern level ($\gamma$) on the effectiveness of an IEA when the reference point is a given constant. The ambition of the pledge is a design parameter, which could be the result of negotiations among participating 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.

### 4.1.1 Impact of the ambition parameter

Figures 3 and 4 illustrate the impact of the ambition of the pledge on the equilibrium solution of the membership and emissions games when the reference point is constant. These results are obtained by unilaterally varying the value of $\alpha$ in the setting described in Figure 1. An increase in the value of $\alpha$ 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.

Figure 3 shows the impact of the ambition parameter on the solution of the membership game and the corresponding steady-state pollution stock. When $\alpha$ is sufficiently high ($\alpha \geq \overline{\alpha}$), no country will join the agreement and $\alpha$ has no impact on the solution, while full participation is attained when $\alpha$ is sufficiently low ($\alpha \leq \underline{\alpha}$).[13] 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 $\alpha$. This illustrates that there is a negative relationship between ambition and participation, which has consequences on environmental quality.

The left panel of Figure 4 shows the impact of $\alpha$ 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 $\alpha$–that is, the emissions of brown countries decrease with the ambition of green countries–but at a much lower rate than that of

[12] 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.

[13] Note that, under full participation, changes in $\alpha$ keep affecting the utility obtained by green countries.

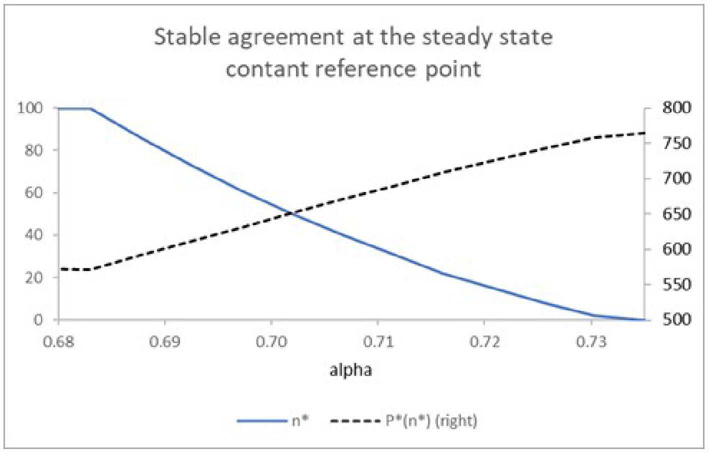


**Figure 3: Left axis: number of green countries in a stable agreement at the steady state. Right axis: pollution stock at the steady state for a stable agreement.** Values are represented as a function of the ambition parameter $\alpha$ for $r_0 = 0.594$ and $\gamma = 0.00065$. Other parameter values are listed in Table 1.

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 $\alpha$. However, green countries have an additional incentive to reduce their emissions due tothe ambition parameter, that is, compliance with the agreement. Note that for $\alpha < \underline{\alpha}$ (full participation), the impact of $\alpha$ on the emissions from (green) countries is negative, but much smaller than for $\alpha \in (\underline{\alpha}, \overline{\alpha})$, since $\alpha$ no longer impacts participation.

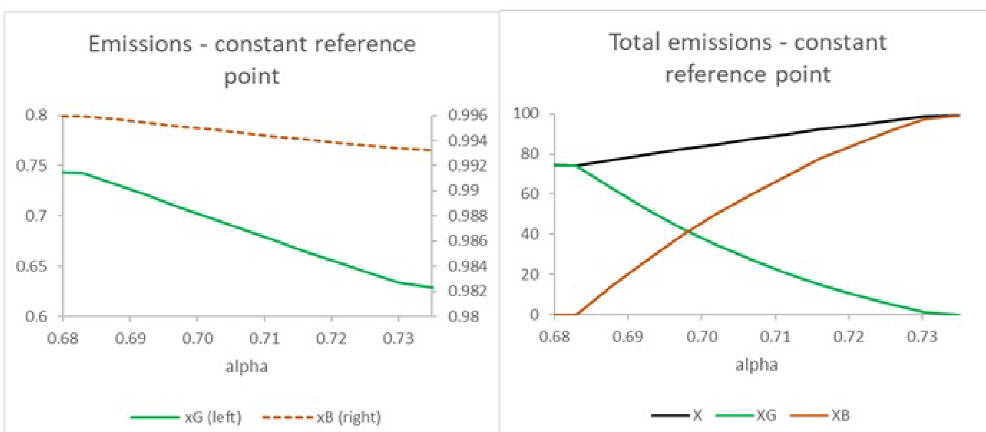


**Figure 4: Left panel: Emissions of green (left axis) and brown (right axis) countries. Right panel: total emissions and total emissions in each group.** Values are reported as a function of the ambition parameter $\alpha$ for $r_0 = 0.594$ and $\gamma = 0.00065$. Other parameter values are listed in Table 1.

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 $\alpha$, as illustrated in the right panel of Figure 4. 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, Figure 5 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 (16)–(17). Since both suffer the same environmental damage cost, 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. Figure 5 shows that both the total discounted equilibrium welfare and the total discounted equilibrium utility at the steady state are decreasing with the ambition parameter $\alpha$. 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.

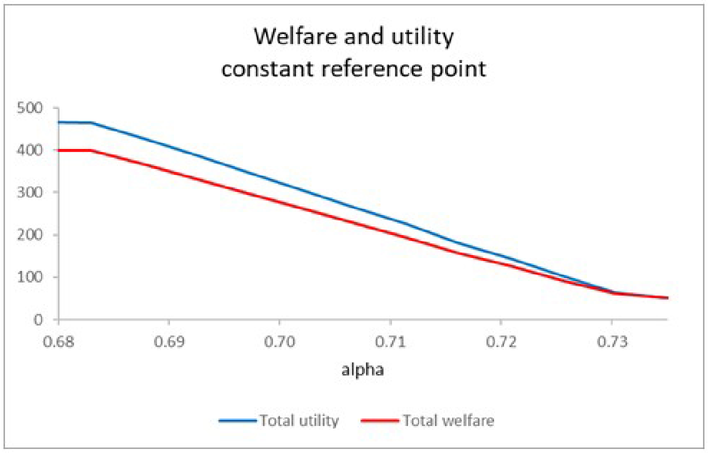


**Figure 5: Total discounted welfare and utility at the steady state as a function of the ambition parameter $\alpha$ for $r_0 = 0.594$ and $\gamma = 0.00065$.** Other parameter values are listed in Table 1.

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 $\alpha$, while the total emissions and pollution stock are increasing with $\alpha$. The overall conclusion of these experiments is that, in the constant-reference-point 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 $\gamma$, again in the setting described in Figure 1. Increasing $\gamma$ increases the weight of self-image and therefore has a direct impact on the value of compliance. Results are reported in Figures 6 to 8.

Figure 6 shows the impact of the level of concern parameter $\gamma$ on the solution of the membership game and steady-state pollution stock. For sufficiently small $\gamma$ ($\gamma \leq \underline{\gamma}$), 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 $\gamma$, 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 Figure 7 shows the impact of $\gamma$ 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 $\gamma$. 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 $\gamma$. 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 $\gamma$. Brown countries' emissions increase with the environmental awareness of green countries because they respond to the stock of pollution, which is decreasing in $\gamma$. In the case of green countries, two opposing effects are at work when $\gamma$ increases: as for brown countries, green countries increase their emissions in response to a lower stock of pollution;

however, an increase in $\gamma$ 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 $\gamma$, and its impact dominates that of the reduction in the pollution stock, so that green countries' emissions are decreasing with $\gamma$. When full participation is reached, the impact of $\gamma$ 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.

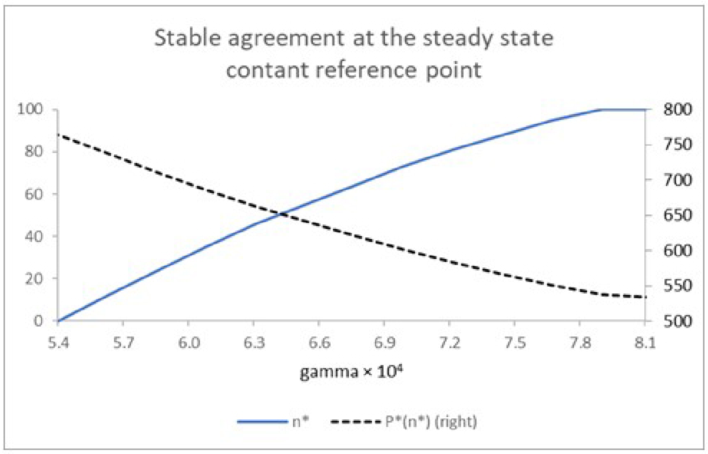


**Figure 6: Left axis: number of green countries in a stable agreement at the steady state. Right axis: pollution stock at the steady state for a stable agreement.** Values are represented as a function of the concern level parameter $\gamma$ for $r_0 = 0.594$ and $\alpha = 0.7$. Other parameter values are listed in Table 1.

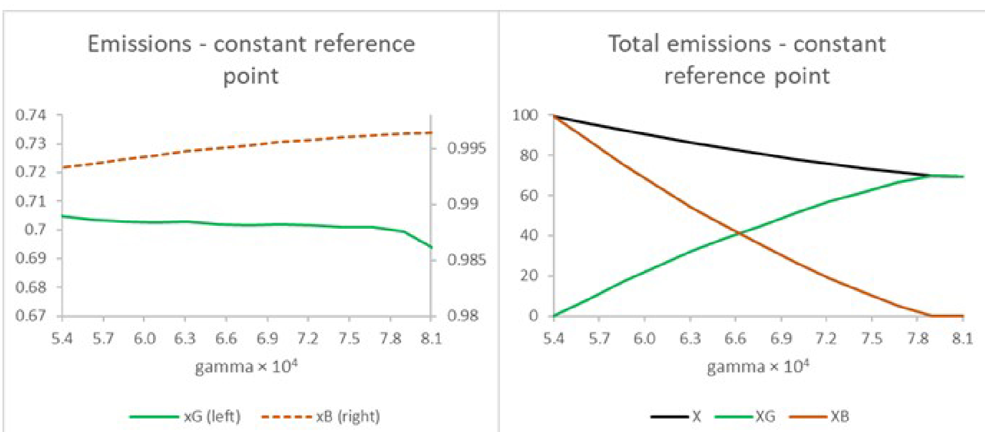


**Figure 7: Left panel: Emissions of green (left axis) and brown (right axis) countries. Right panel: Total emissions and total emissions in each group.** Values are represented as a function of the concern level parameter $\gamma$ for $r_0 = 0.594$ and $\alpha = 0.7$. Other parameter values are listed in Table 1.

The impact of $\gamma$ on the total emissions is illustrated in the right panel of Figure 7, showing that participation in the IEA drives the result. While brown countries' individual emissions are increasing with $\gamma$, their total contribution is decreasing because the number of brown countries is decreasing with $\gamma$. The reverse is true for green countries: as $\gamma$ 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 $\gamma$.

Finally, Figure 8 shows that both the total discounted welfare and the total discounted utility are increasing with the concern level parameter $\gamma$.

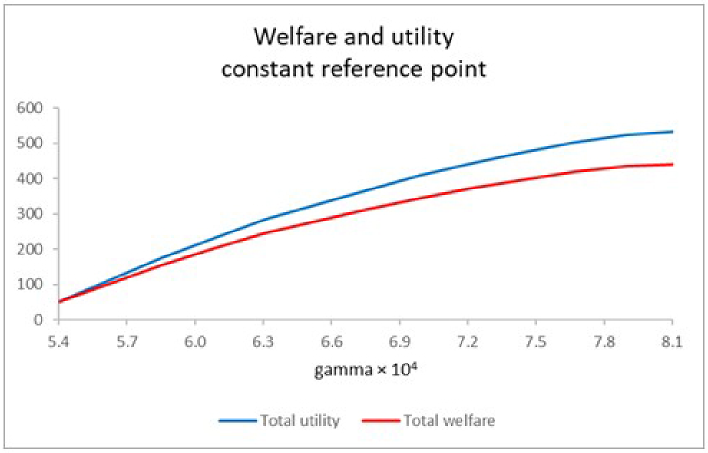


**Figure 8: Total discounted welfare and utility at the steady state, as a function of the concern level parameter $\gamma$ for $r_0 = 0.594$ and $\alpha = 0.7$.** Other parameter values are listed in Table 1.

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 reference point

In the second set of experiments, the reference point depends on the current pollution stock ($r_1 < 0$). Choosing a state-dependent reference point is equivalent to choosing a reference strategy, here, linear in the pollution stock. 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 emissions level is halfway between the two corresponding levels for all $P$.

Figure 9 shows that stable agreements can be achieved when the target is decreasing with the pollution stock; for the set of parameter values used in Figure 9, the solution of the membership game is $n^* = 39$ and the correspondent steady-state pollution stock sits at $P^*(39) = 646$.

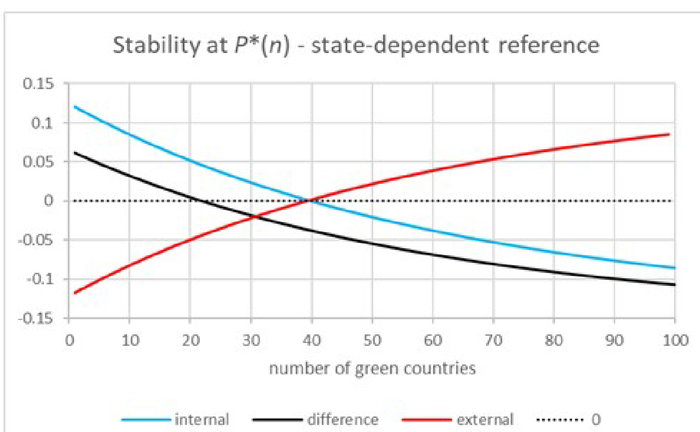


**Figure 9: Stability conditions at $P^*(n)$ as a function of the number of green (signatory) countries $n$.** Other parameter valuesThe self-image function parameter values are $r_0 = 0.887$, $r_1 = -0.0002$, $\alpha = 0.95$, $\gamma = 0.00065$. are listed in Table 1.

Figure 10 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 at $n = 39$.

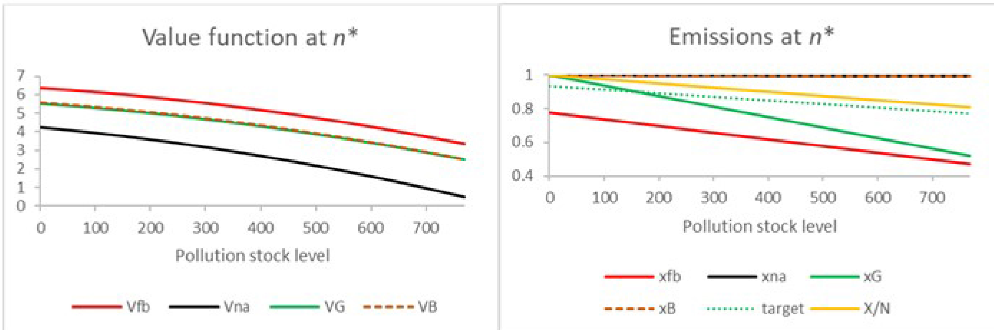


**Figure 10: Equilibrium value function (left panel) and emissions (right panel) of green and brown countries as a function of the pollution stock for $n = 39$, compared to the no-agreement and the first-best solutions.** Self-image parameter values are $r_0 = 0.887$, $r_1 = -0.0002$, $\alpha = 0.95$, $\gamma = 0.00065$. Other parameter values are from Table 1.

Note that with a state-dependent reference point, the target is no longer constant but is decreasing with the pollution stock; because the equilibrium strategies are linear in $P$, it is easier for green countries to track the 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 reference point is qualitatively similar to the results obtained with a constant reference point: 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 close to 0. As in the constant-reference-point case, it is possible to find combinations of the ambition of the pledge and weight of the self-image parameters such that an agreement with partial participation can outweigh the first-best solution in terms of welfare for all countries and for all levels of the pollution stock.[14]

Appendix A.5.3 presents the results of sensitivity analyses performed with respect to the ambition of the pledge $\alpha$ and the level of concern $\gamma$ for the set of parameters used in Figure 9. Our numerical investigations with various sets of parameters show that the main difference with respect to the constant-reference-point case is that players' strategies and utilities are significantly more sensitive to variations in the self-image parameter values.

Qualitatively, the impact of unilateral variations in the ambition parameter $\alpha$ is similar to what is reported for the constant-target scenario; however, sensitivity analyses of the individual emissions with respect to the concern parameter $\gamma$ show that, contrary to what happens when the target is constant, both brown and green countries' emissions are increasing in $\gamma$ (see Figure A.8). This result is due to the predominant impact of the decrease in pollution stock over the increase of $\gamma$. Nevertheless, as in the constant-reference-point 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 that that of brown countries, the total emissions decrease with $\gamma$ as the number of participants in the agreement increases.

## 4.3 Trajectories over time

Recall that all the results and the analysis provided in the previous sections are done at the steady state, where the number of signatories and the stock of pollution are stable. In an open-membership

[14] With the same reference point, this happens for instance at $\alpha = 0.95$ and $\gamma = 0.0007$.

setting, both the number of signatory countries and the level of pollution stock can change from period to period. In this section, we provide examples of time trajectories for the pollution stock and for the number of signatories of an IEA.

The system evolves in discrete 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 Equation (1). As to 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,

$$n_t = n^*(P_t).$$

This scenario corresponds to the assumption made in Rubio & Ulph (2007). In this setting, decision dates can be interpreted as negotiation dates at which the membership game is played and the set of signatory countries is reshuffled. In other words, at each decision date, countries observe the level of the pollution stock and make their membership decisions, so that the resulting number of signatory countries solves conditions (16) and (17). 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 every time 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 its solution 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** In Scenario S3, 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, in and out of the agreement. More precisely,

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

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 Scenario S2, the participation level in the previous period has an impact on that of the next period.

Figure 11 shows examples of trajectories under the three scenarios for the parameter values used in the example in Figure 1, where the steady state is 54 signatory countries and a pollution stock of 644. In Figure 11, the three 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. We observe from these examples that all three scenarios produce trajectories that converge to the steady state. The main difference between these trajectories is that the ones generated under Scenario 1 take a direct path towards the steady state,[15] while the trajectories generated by the dynamics under Scenarios 2 and 3 tend to overshoot (*resp. undershoot*) both the stable size and the pollution stock in the left (*resp. right*) panel. This occurs because, in Scenarios 2 and 3, changes are gradual, whereas in Scenario 1, the variation in the number of countries from one period to the next can be significant.

[15] Clearly, the path under S1 coincides with the function $n^*(P)$. 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 most of the feasible range for the pollution stock.

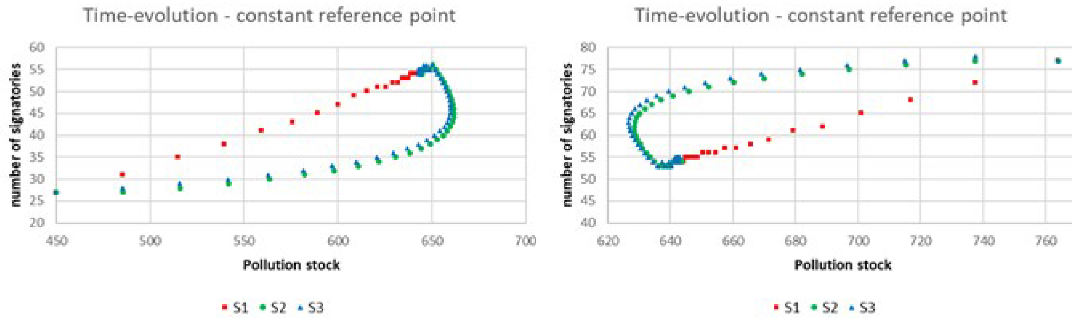


**Figure 11: Trajectories for the pollution stock and the number of signatories over 50 time periods under Scenarios S1, S2, and S3. Initial conditions are $(n_0, P_0) = (450, 27)$ in the left panel and $(764, 77)$ in the right panel.** The self-image function parameter values are $r_0 = 0.594$, $\alpha = 0.7$, $\gamma = 6.5\ \times 10^{-4}$.Other parameter values are listed in Table 1.

However, there is no reason in Scenarios S2 and S3 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. Figure 12 compares the trajectories under the three scenarios starting from the business-as-usual pollution stock ($P_0 = 764$), and from $n_0 = 1$ (left panel) and $n_0 = 90$ (right panel) for Scenarios S2 and S3. Note that the trajectory under Scenario S2 has not yet converged to the steady state after 50 time steps when $n_0 = 1$.

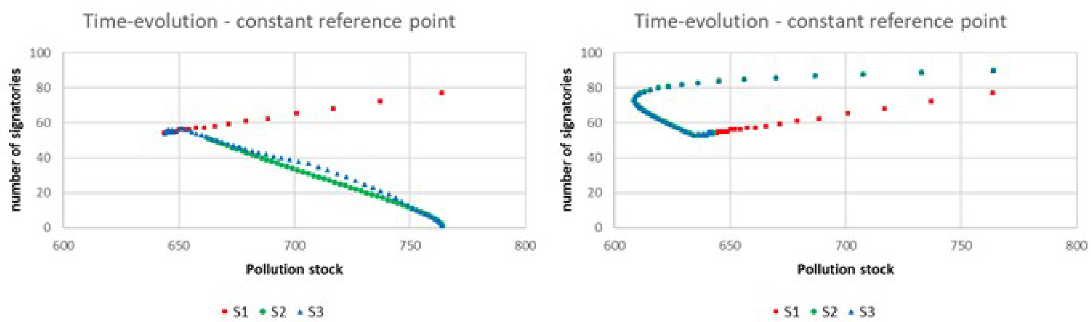


**Figure 12: Trajectories for the pollution stock and the number of signatories over 50 time periods under Scenarios S1, S2, and S3. Initial pollution stock is $P_0 = 764$. In the left panel, $n_0 = 1$ and in the right panel, $n_0 = 90$ for Scenarios S2 and S3.** The self-image function parameter values are $r_0 = 0.594$, $\alpha = 0.7$, $\gamma = 6.5\ \times 10^{-4}$. Other parameter values are listed in Table 1.

# 5 Conclusion

In this paper we elaborated a (dynamic) model that contributes to the literature on the size of a stable IEA aimed at reducing pollution emissions.

Inspired by the approaching deadline of COP26 and by what it entails for countries party to the Paris Agreement (2015), the model adopts three of its specific features, namely: transparency, ambition of the pledge, and freedom of action. This has been operationalized by defining signatory countries as countries that value a self-image arising from their commitment to the agreement, and by assuming that all countries choose their emissions strategies independently.

Our investigation has shown 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. It is even possible, when the weight of the self-image is sufficiently high, that stable agreements with partial participation outweigh the first-best solution.

To better understand the role of self-image in the stability and efficiency of agreements, we analyzed 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. A sensitivity analysis with respect to the ambition of the pledge showed that there is a trade-off between ambition and participation, which has negative environmental and economic consequences. On the other hand, a sensitivity analysis with respect to countries' awareness of the risks and the dangers of the environmental problem at stake has shown that there is a fortifying relationship between participation and countries' concern, with positive environmental and economic consequences. Those results are robust to the assumptions about the type of reference point (constant or indexed on the current pollution stock level). It would be interesting to investigate if those results are also robust to other types of reference points, as would be the case, for example, for "status concerned" countries. Furthermore, the choice of the ambition of the pledge, which is a design parameter, could become part of the model.

The result of our investigations can be summarized by noting 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 collective action is needed to tackle global problems like climate change. As countries approach COP26, in light of the current debate about the level of ambition of their new NDCs, we submit that a large collective action, induced by a modest level of ambition, is more effective than a few countries committing highly curbing their emissions.

# Appendix

## A.1 Proof of Proposition 1

Assume that the value function takes the form

$$V(P) = b_2P^2 + b_1P + b_0.$$

The value function then satisfies

$$V(P) = \max_x \left\{c_1x^2 + c_2x + c_3Px + c_4P^2 + c_5P + \beta\left(b_2\left(\delta P + O + x\right)^2 + b_1\left(\delta P + O + x\right) + b_0\right)\right\}$$

where $O$ denotes the emissions of other countries.

First-order conditions yield

$$x = -\frac{1}{2}\frac{c_2 + b_1\beta + Pc_3 + 2Xb_2\beta + 2Pb_2\beta\delta}{c_1}$$

where $X = O + x$ is the total emissions by all countries. This is the best response of the country, provided that the second-order conditions are satisfied, that is,

$$2\left(c_1 + b_2\beta\right) < 0.$$

Therefore, under our assumption about the form of the value function, if the other countries use a strategy that is linear in $P$, then the best response of a single country is also linear in $P$, and $X$ is therefore linear in $P$. Let

$$\begin{aligned} X =& Q_0 + Q_1P \\ x =& -\frac{1}{2}\frac{c_2 + \beta\left(b_1 + 2Q_0b_2\right) + P\left(c_3 + 2b_2\beta\left(Q_1 + \delta\right)\right)}{c_1}. \end{aligned}$$

To ease notation, define

$$M_1 \equiv Q_1 + \delta$$

$$M_2 = b_1 + 2Q_0 b_2.$$

Replacing $x$ and $X$ in the expression of the value function yields

$$\begin{aligned}
V(P) =& P^2 \left( c_4 - \frac{1}{4c_1} c_3^2 + \beta M_1^2 b_2 \frac{c_1 + \beta b_2}{c_1} \right) \\
&+ P \left( c_5 + \frac{-c_2 c_3 + 2\beta^2 M_1 M_2 b_2}{2c_1} + \beta M_1 M_2 \right) \\
&+ \left( \frac{M_2^2 \beta^2 - c_2^2}{4c_1} + \beta \left( b_0 + Q_0 M_2 \right) \right).
\end{aligned}$$

This shows that the value function is quadratic in $P$, provided we can find constants $b_2$, $b_1$, and $b_0$ satisfying

$$\begin{aligned}
b_2 =& c_4 - \frac{1}{4c_1} c_3^2 + \beta M_1^2 b_2 \frac{c_1 + \beta b_2}{c_1} \\
b_1 =& c_5 + \frac{-c_2 c_3 + 2\beta^2 M_1 M_2 b_2}{2c_1} + \beta M_1 M_2 \\
b_0 =& \frac{M_2^2 \beta^2 - c_2^2}{4c_1} + \beta \left( b_0 + Q_0 M_2 \right).
\end{aligned}$$

## A.2 Proof of Proposition 2

The equilibrium solution satisfies

$$\begin{aligned}
x^B =& -\frac{1}{2} \frac{c_2^B + \beta \left( b_1^B + 2Q_0 b_2^B \right) + P \left( c_3^B + 2b_2^B \beta \left( Q_1 + \delta \right) \right)}{c_1^B} \\
\equiv& q_0^B + q_1^B P \\
x^G =& -\frac{1}{2} \frac{c_2^G + \beta \left( b_1^G + 2Q_0 b_2^G \right) + P \left( c_3^G + 2b_2^G \beta \left( Q_1 + \delta \right) \right)}{c_1^G} \\
\equiv& q_0^G + q_1^G P.
\end{aligned}$$

Using $Q_0 = nq_0^G + mq_0^B$ and $Q_1 = nq_1^G + mq_1^B$ yields two linear systems of equations

$$\begin{aligned}
q_0^B =& -\frac{1}{2} \frac{c_2^B + \beta \left( b_1^B + 2b_2^B \left( nq_0^G + mq_0^B \right) \right)}{c_1^B} \\
q_0^G =& -\frac{1}{2} \frac{c_2^G + \beta \left( b_1^G + 2b_2^G \left( nq_0^G + mq_0^B \right) \right)}{c_1^G} \\
q_1^B =& -\frac{1}{2} \frac{c_3^B + 2b_2^B \beta \left( nq_1^G + mq_1^B + \delta \right)}{c_1^B} \\
q_1^G =& -\frac{1}{2} \frac{c_3^G + 2b_2^G \beta \left( nq_1^G + mq_1^B + \delta \right)}{c_1^G}
\end{aligned}$$

and the solution is readily obtained as

$$\begin{aligned}
q_0^B =& \frac{b_2^B n\beta \left( c_2^G + b_1^G \beta \right) - \left( c_1^G + b_2^G n\beta \right) \left( c_2^B + b_1^B \beta \right)}{2 \left( \Phi\beta + c_1^B c_1^G \right)} \\
q_0^G =& \frac{b_2^G m\beta \left( c_2^B + b_1^B \beta \right) - \left( c_1^B + b_2^B m\beta \right) \left( c_2^G + b_1^G \beta \right)}{2 \left( \Phi\beta + c_1^B c_1^G \right)} \\
Q_0 =& -\frac{1}{2} \frac{mc_1^G \left( c_2^B + \beta b_1^B \right) + nc_1^B \left( c_2^G + \beta b_1^G \right)}{\Phi\beta + c_1^B c_1^G}
\end{aligned}$$

$$
\begin{aligned}
q_1^B =& \frac{b_2^B n\beta\left(c_3^G + 2b_2^G\beta\delta\right) - \left(c_1^G + b_2^G n\beta\right)\left(c_3^B + 2b_2^B\beta\delta\right)}{2\left(\Phi\beta + c_1^B c_1^G\right)} \\
q_1^G =& \frac{b_2^G m\beta\left(c_3^B + 2b_2^B\beta\delta\right) - \left(c_1^B + b_2^B m\beta\right)\left(2b_2^G\beta\delta + c_3^G\right)}{2\left(\Phi\beta + c_1^B c_1^G\right)} \\
Q_1 =& -\frac{1}{2}\frac{mc_1^G c_3^B + nc_1^B c_3^G + 2\Phi\beta\delta}{\Phi\beta + c_1^B c_1^G}
\end{aligned}
$$

where

$$
\Phi \equiv mb_2^B c_1^G + nb_2^G c_1^B .
$$

Using

$$
\begin{aligned}
c_1^G =& c_1^B = -\frac{1}{2} \\
c_2^G =& c_2^B = 1 \\
c_3^G =& -\alpha\gamma \ ; \ c_3^B = 0 \\
c_4^G =& \frac{1}{\alpha}\alpha\gamma r_1 - d_2 \ ; \ c_4^G = -d_2
\end{aligned}
$$

yields the result.

## A.3 Value function parameters as a function of $b_2^G$, $b_2^B$, and $Q_1$

For given values of $b_2^G$, $b_2^B$, and $\Phi$, $Q_1$ and $M_1$ are constants and the global value $Q_0$ is linear in $b_1^G$ and $b_1^B$ :

$$
\begin{aligned}
Q_0 =& -\frac{1}{2}\frac{mc_1^G\left(c_2^B + \beta b_1^B\right) + nc_1^B\left(c_2^G + \beta b_1^G\right)}{\Phi\beta + c_1^B c_1^G} \\
=& \left(-\frac{1}{2}\frac{mc_1^G\left(c_2^B + \beta b_1^B\right) + nc_1^B\left(c_2^G + \beta b_1^G\right)}{D}\right).
\end{aligned}
$$

As a result, conditions (7) for $B$ and $G$ players is a system of two linear equations

$$
\begin{aligned}
b_1^B =& c_5^B + \frac{-c_2^B c_3^B + 2\beta^2 M_1\left(b_1^B + 2Q_0 b_2^B\right)b_2^B}{2c_1^B} + \beta M_1\left(b_1^B + 2Q_0 b_2^B\right) \\
b_1^G =& c_5^G + \frac{-c_2^G c_3^G + 2\beta^2 M_1\left(b_1^G + 2Q_0 b_2^G\right)b_2^G}{2c_1^G} + \beta M_1\left(b_1^G + 2Q_0 b_2^G\right)
\end{aligned}
$$

and the solution is readily obtained analytically:

$$
\begin{aligned}
b_1^B =& \frac{A^B n\beta b_2^B c_1^B\left(H^G c_1^G + A^G F b_2^G\right) + c_1^G\left(A^G\left(c_1^B + m\beta b_2^B\right) - 1\right)\left(H^B c_1^B + A^B F b_2^B\right)}{c_1^G c_1^B\left(\left(A^G K^B - 1\right)\left(A^B K^G - 1\right) - A^B A^G mn\beta^2 b_2^B b_2^G\right)} \\
b_1^G =& \frac{A^G m\beta b_2^G c_1^G\left(H^B c_1^B + A^B F b_2^B\right) + c_1^B\left(A^B\left(c_1^G + n\beta b_2^G\right) - 1\right)\left(H^G c_1^G + A^G F b_2^G\right)}{c_1^B c_1^G\left(\left(A^G K^B - 1\right)\left(A^B K^G - 1\right) - A^B A^G mn\beta^2 b_2^B b_2^G\right)}
\end{aligned}
$$

where

$$
\begin{aligned}
D =& \Phi\beta + c_1^B c_1^G \\
A^B =& \frac{\beta M_1\left(c_1^B + \beta b_2^B\right)}{D} ; A^G = \frac{\beta M_1\left(c_1^G + \beta b_2^G\right)}{D} \\
F =& mc_1^G c_2^B + nc_1^B c_2^G
\end{aligned}
$$

$$
\begin{aligned}
H^B &= -c_5^B + \frac{1}{2c_1^B} c_2^B c_3^B ; H^G = -c_5^G + \frac{1}{2c_1^G} c_2^G c_3^G \\
K^B &= \left(c_1^B + m\beta b_2^B\right) ; K^G = \left(c_1^G + n\beta b_2^G\right).
\end{aligned}
$$

Finally, Conditions (8) for $B$ and $G$ players yield

$$
\begin{aligned}
b_0^B (1-\beta) =& \frac{\left(b_1^B + 2Q_0 b_2^B\right)^2 \beta^2 - \left(c_2^B\right)^2}{4c_1^B} + \beta Q_0 \left(b_1^B + 2Q_0 b_2^B\right) \\
b_0^G (1-\beta) =& \frac{\left(b_1^G + 2Q_0 b_2^G\right)^2 \beta^2 - \left(c_2^G\right)^2}{4c_1^G} + \beta Q_0 \left(b_1^G + 2Q_0 b_2^G\right).
\end{aligned}
$$

Using

$$
\begin{aligned}
c_1^G =& c_1^B = -\frac{1}{2} \\
c_2^G =& c_2^B = 1 \\
c_3^G =& -\alpha\gamma \ ; \ c_3^B = 0 \\
c_4^G =& \frac{1}{\alpha}\alpha\gamma r_1 - d_2 \ ; \ c_4^B = -d_2 \\
c_5^G =& \frac{1}{\alpha}\alpha\gamma r_0 - d_1 \ ; \ c_5^B = -d_1,
\end{aligned}
$$

the equilibrium value function parameters are given by

$$
\begin{aligned}
b_1^B =& \frac{A^B n\beta b_2^B \left(\alpha\gamma\frac{\alpha - r_0}{\alpha} + d_1 + A^G N b_2^G\right) + \frac{1}{2}\left(d_1 + A^B N b_2^B\right)\left(A^G \left(2m\beta b_2^B - 1\right) - 2\right)}{\left(A^G K^B - 1\right)\left(A^B K^G - 1\right) - A^B A^G mn\beta^2 b_2^B b_2^G} \\
b_1^G =& \frac{A^G m\beta b_2^G \left(d_1 + A^B N b_2^B\right) + \frac{1}{2}\left(\alpha\gamma\frac{\alpha - r_0}{\alpha} + d_1 + A^G N b_2^G\right)\left(A^B \left(2n\beta b_2^G - 1\right) - 2\right)}{\left(A^G K^B - 1\right)\left(A^B K^G - 1\right) - A^B A^G mn\beta^2 b_2^B b_2^G}
\end{aligned}
$$

$$
\begin{aligned}
b_0^B (1-\beta) =& \frac{1 - \left(b_1^B + 2Q_0 b_2^B\right)^2 \beta^2}{2} + \beta Q_0 \left(b_1^B + 2Q_0 b_2^B\right) \\
b_0^G (1-\beta) =& \frac{1 - \left(b_1^G + 2Q_0 b_2^G\right)^2 \beta^2}{2} + \beta Q_0 \left(b_1^G + 2Q_0 b_2^G\right)
\end{aligned}
$$

$$
\begin{aligned}
A^B =& 2\beta M_1 \frac{2\beta b_2^B - 1}{4\Phi\beta + 1} ; A^G = 2\beta M_1 \frac{2\beta b_2^G - 1}{4\Phi\beta + 1} \\
K^B =& m\beta b_2^B - \frac{1}{2} ; K^G = n\beta b_2^G - \frac{1}{2}.
\end{aligned}
$$

## A.4 Cooperative equilibrium

Assume that the value function takes the form

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

The value function then satisfies

$$
V^C(P) = \max_x \left\{ -\frac{1}{2} x (x-2) - d_2 P^2 - d_1 P + \beta \left( b_2^C (\delta P + N x)^2 + b_1^C (\delta P + N x) + b_0^C \right) \right\}.
$$

First-order conditions yield

$$
x = \frac{1 + N\beta \left(b_1^C + 2P\delta b_2^C\right)}{1 - 2N^2 \beta b_2^C}.
$$

This emissions strategy is linear in $P$ and maximizes the total welfare, provided that the second-order conditions are satisfied, that is,

$$1 - 2N^2\beta b_2^C > 0.$$

Replacing $x$ in the expression of the value function yields

$$\begin{aligned} V(P) =& P^2 \left( \frac{d_2 - \beta b_2^C \left(\delta^2 + 2N^2 d_2\right)}{2N^2\beta b_2^C - 1} \right) \\ &+ P \left( \frac{d_1 - \beta \left(\delta b_1^C + 2N b_2^C \left(\delta + N d_1\right)\right)}{2N^2\beta b_2^C - 1} \right) \\ &+ \frac{1}{2} \frac{\beta \left(-2b_0^C + N \left(-2b_1^C + N\beta \left(4b_0^C b_2^C - b_1^{2C}\right)\right)\right) - 1}{2N^2\beta b_2^C - 1}. \end{aligned}$$

This shows that the value function is indeed quadratic in $P$, with

$$\begin{aligned} b_2^C =& \frac{1 - \beta\left(\delta^2 + 2N^2 d_2\right) - \sqrt{\left(\beta\delta^2 + 2N^2\beta d_2 - 1\right)^2 + 8N^2\beta d_2}}{4N^2\beta} \\ b_1^C =& \frac{2N\beta b_2^C \left(\delta + N d_1\right) - d_1}{1 - \beta\left(\delta + 2N^2 b_2^C\right)} \\ b_0^C =& \frac{1 + N b_1^C \beta \left(N\beta b_1^C + 2\right)}{2\left(1-\beta\right)\left(1 - 2N^2\beta b_2^C\right)}. \end{aligned}$$

## A.5 Additional figures

### A.5.1 Low pollution stock

Figure A.1 is an enlargement of Figure 2 corresponding to very small values of the pollution stock.

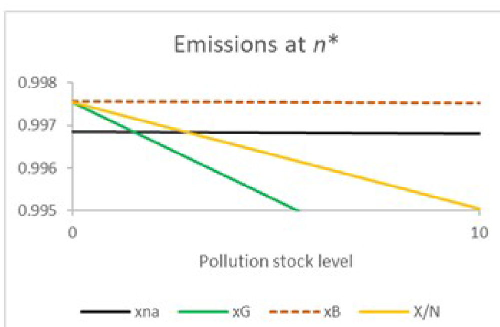


**Figure A.1: Emissions of green and brown countries and their weighted average at $n = 54$ for small values of the pollution stock, compared to the no-agreement solution.** Self-image parameter values are $r_0 = 0.594$, $\alpha = 0.7$, $\gamma = 6.5\ \times 10^{-4}$. Other parameter values are from Table 1.

### A.5.2 Example with a highly weighted self-image

Figure A.2 shows the solution of the membership game and A.3 illustrates the value functions and the emissions at $n^*$ as a function of the current pollution stock. This example is obtained with an ambition parameter ($\alpha = 0.84$) that produces a more stringent target $T = 0.707$ and a concern parameter ($\gamma = 0.0011$) that is twice as high as in the example illustrated in Figures 1–2.

The agreement is stable for the same number of signatory countries ($n^* = 54$), but it results in a lower steady-state pollution stock ($P^*(54) = 528$). At the steady state, 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. Furthermore, Figure A.3 shows that for that set of parameter values, the long-term equilibrium utilities of both green and brown countries are higher than the first-best solution, for all possible levels of the pollution stock.

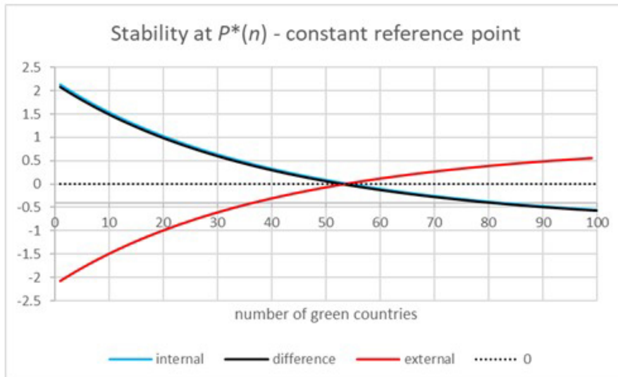


**Figure A.2: Stability conditions at $P^*(n)$ as a function of the number of green (signatory) countries $n$.** The self-image function parameter values are $r_0 = 0.594$, $\alpha = 0.837$, $\gamma = 13 \times 10^{-4}$. Other parameter values are listed in Table 1.

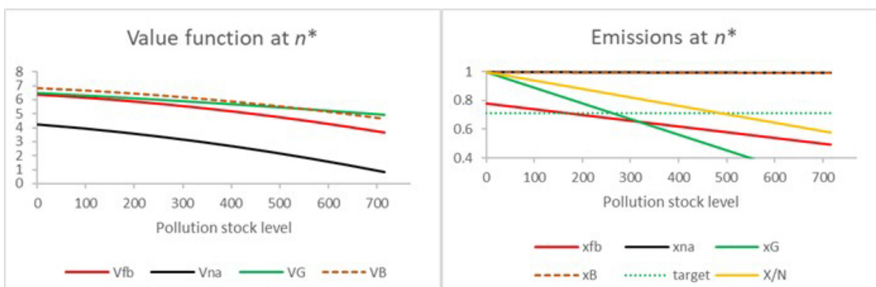


**Figure A.3: Equilibrium value function (left panel) and emissions (right panel) of green and brown countries as a function of the pollution stock for $n = 54$, compared to the first-best and the no-agreement solutions.** Self-image parameter values are $r_0 = 0.594$, $\alpha = 0.837$, $\gamma = 13 \times 10^{-4}$. Other parameter values are from Table 1.

### A.5.3 Sensitivity analysis, state-dependent reference point

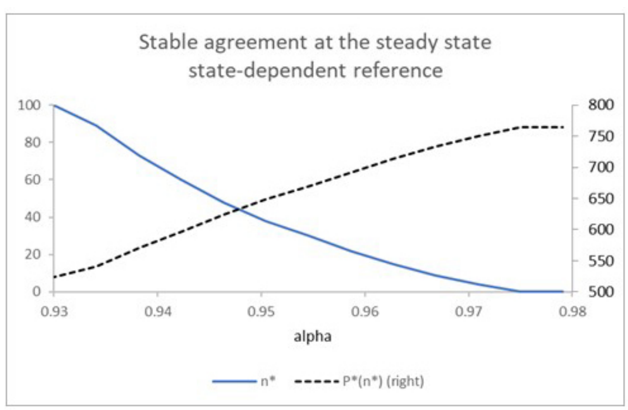


**Figure A.4: Left axis: number of green countries in a stable agreement at the steady state. Right axis: pollution stock at the steady state for a stable agreement.** Values are represented as a function of the ambition parameter $\alpha$ for $r_0 = 0.887$, $r_1 = -0.0002$, and $\gamma = 0.00065$. Other parameter values are listed in Table 1.

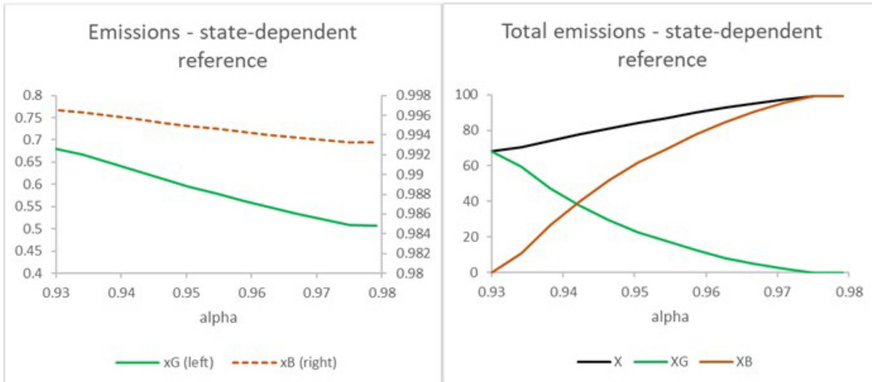


**Figure A.5: Left panel: emissions of green (left axis) and brown (right axis) countries. Right panel: total emissions and total emissions in each group.** Values are represented as a function of the ambition parameter $\alpha$ for $r_0 = 0.887$, $r_1 = -0.0002$, and $\gamma = 0.00065$. Other parameter values are listed in Table 1.

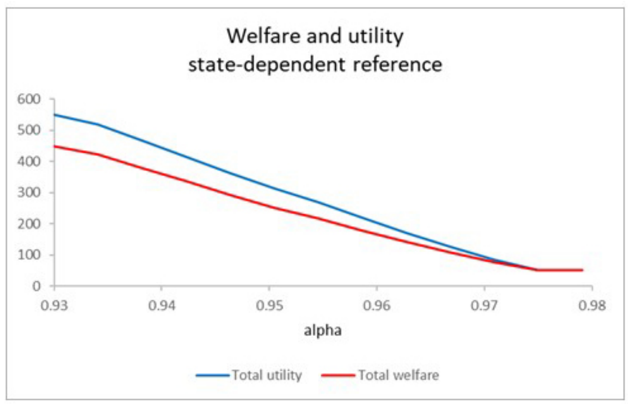


**Figure A.6: Total discounted welfare and utility at the steady state as a function of the ambition parameter $\alpha$ for $r_0 = 0.887$, $r_1 = -0.0002$, and $\gamma = 0.00065$.** Other parameter values are listed in Table 1.

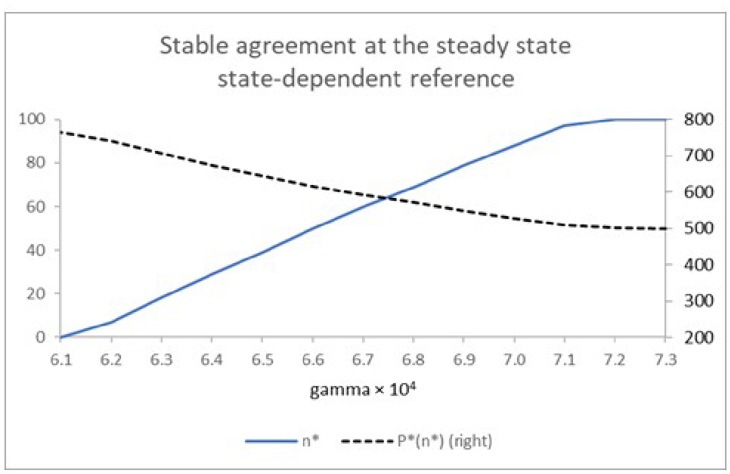


**Figure A.7: Left axis: number of green countries in a stable agreement at the steady state. Right axis: pollution stock at the steady state for a stable agreement.** Values are represented as a function of the weight parameter $\gamma$ for $r_0 = 0.887$, $r_1 = -0.0002$, and $\alpha = 0.95$. Other parameter values are listed in Table 1.

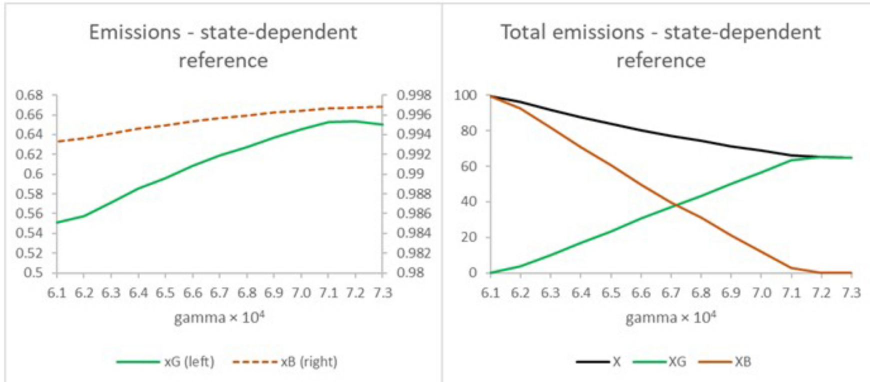


**Figure A.8: Left panel: emissions of green (left axis) and brown (right axis) countries. Right panel: total emissions and total emissions in each group.** Values are represented as a function of the concern parameter $\gamma$ for $r_0 = 0.887$, $r_1 = -0.0002$, and $\alpha = 0.95$. Other parameter values are listed in Table 1.

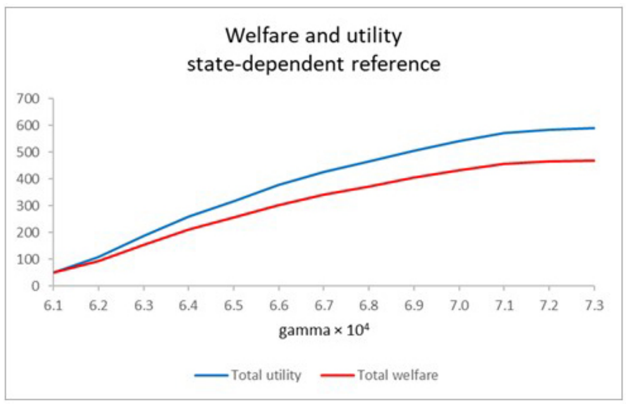


**Figure A.9: Total discounted welfare and utility at the steady state as a function of the concern parameter $\gamma$ for $r_0 = 0.887$, $r_1 = -0.0002$, and $\alpha = 0.95$.** Other parameter values are listed in Table 1.

# References


[1] d'Aspremont, C., Jacquemin, A., Gabszewicz, J.J. & Weymark, J.A. (1983). On the stability of collusive price leadership. Canadian Journal of economics pp. 17–25.

[2] Bahn, O., Breton, M., Sbragia, L. & Zaccour, G. (2009). Stability of international environmental agreements: an illustration with asymmetrical countries. International Transactions in Operational Research 16(3): 307–324.

[3] Barrett, S. (1994). Self-enforcing international environmental agreements. Oxford Economic Papers 46: 878–894.

[4] Barrett, S. (1997). The strategy of trade sanctions in international environmental agreements. Resource and Energy Economics 19(4): 345–361.

[5] Barrett, S. (2003). Environment and statecraft: The strategy of environmental treaty-making. Oxford University Press, Oxford.

[6] Benchekroun, H. & van Long, N. (2016). Status concern and the exploitation of common pool renewable resources. Ecological Economics 125: 70–82.

[7] 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.

[8] Brekke, K.A., Kverndokk, S. & Nyborgb, K. (2003). An economic model of moral motivation. Journal of Public Economics 87(9–10): 1967–1983.

[9] Breton, M. & Garrab, S. (2014). Evolutionary farsightedness in international environmental agreements. International Transactions in Operational Research 21(1): 21–39.

[10] Breton, M., Sbragia, L. & Zaccour, G. (2010). A dynamic model for international environmental agreements. Environmental and Resource economics 45(1): 25–48.

[11] Card, D., Mas, A., Moretti, E. & Saez, E. (2012). Inequality at Work: The Effect of Peer Salaries on Job Satisfaction. American Economic Review 102(6): 2981–3003.

[12] Carraro, C., Marchiori, C. & Oreffice, S. (2009). Endogenous Minimum Participation in International Environmental Treaties. Environmental and Resource Economics 42: 411–425.

[13] Carraro, C. & Siniscalco, D. (1993). Strategies for international protection of the environment. Journal of Public Economics 52: 309–328.

[14] 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.

[15] Colombo, L. & Labrecciosa, P. (2018). Consumer surplus-enhancing cooperation in a natural resource oligopoly. Journal of Environmental Economics and Management 92: 185–193.

[16] Eyckmans, J. & Kverndokk, S. (2010). Moral concerns on tradable pollution permits in international environmental agreements. Ecological Economics 69: 1814–1823.

[17] Fehr, E. & Schmidt, K.M. (1999). A Theory of Fairness, Competition, and Cooperation. The Quarterly Journal of Economics 114(3): 817–868.

[18] Finus, M, & Rübbelke, D.T.G (2013). Public Good Provision and Ancillary Benefits: The Case of Climate Agreements. Environmental and Resource Economics 56: 211–226.

[19] Frank, R.H. (2005). Positional Externalities Cause Large and Preventable Welfare Losses. The American Economic Review 95(2): 137–141.

[20] Grüning, C. & Peters, W. (2010). Can justice and fairness enlarge international environmental agreements? Games 1(2): 137–158.

[21] Hoel, M.(1992). International environment conventions: the case of uniform reductions of emissions. Environmetal and Resource Economics 2(2): 141–159.

[22] Hoel, M. & Schneider, K. (1997). Incentives to participate in an international environmental agreement. Environmental and Resource Economics 9: 153–170.

[23] Lange, A. (2006). The Impact of Equity Preferences on the Stability of International Environmental Agreements. Environmental and Resource Economics 34: 247–67.

[24] Lange A. & Vogt, C. (2003). Cooperation in international environmental negotiations due to a preference for equity. Journal of Public Economics 87: 2049–2067.

[25] van Long, N. (2020). Warm glow and the transmission of pro-socialness across generations. Journal of Public Economic Theory 22(2): 371–387.

[26] van Long, N. (2021). Dynamic Games of Common-Property Resource Exploitation When Self-image Matters. Dynamic Analysis in Complex Economic Environments pp. 81–107.

[27] Marchiori, C., Dietz, S. & Tavoni, A. (2017). Domestic politics and the formation of international environmental agreements. Journal of Environmental Economics and Management 81: 115–131.

[28] Petrakis, E. & Xepapadeas, A. (1996). Environmental consciousness and moral hazard in international agreements to protect the environment. Journal of Public Economics 60: 95–110.

[29] Rubio, S.J. & Ulph, A. (2007). An infinite-horizon model of dynamic membership of international environmental agreements. Journal of Environmental Economics and Management 54(3): 296–310.

[30] van der Pol, T. Weikard, H.T. & van Ierland, E. (2012). Can altruism stabilise international climate agreements? Ecological Economics 81: 112–120.

[31] Weikard, H.P., Wangler, L. & Freytag, A. (2015). Minimum participation rules with heterogeneous countries. Environmental and Resource Economics 62(4): 711–727.

[32] Wirl, F. (2011). Global Warming with Green and Brown Consumers. The Scandinavian Journal of Economics 113(4): 866–884.

[33] de Zeeuw, A. (2008). Dynamic effects on the stability of international environmental agreements. Journal of environmental economics and management 55(2): 163–174.