2020

03

Working Paper

INSTITUTO DE POLÍTICAS Y BIENES PÚBLICOS (IPP)

PLEDGE AND IMPLEMENT BARGAINING IN THE PARIS AGREEMENT ON CLIMATE CHANGE

Alejandro Caparrós

INSTITUTO DE POLÍTICAS Y BIENES PÚBICOS – CSIC

Instituto de Políticas y Bienes Públicos
Consejo Superior de Investigaciones Científicas
C/ Albasanz, 26-28
28037 Madrid (España)

Tel: +34 91 6022300
Fax: +34 91 3045710

http://www.ipp.csic.es

How to quote or cite this document:

Caparrós, A. (2020). Pledge and implement bargaining in the Paris Agreement on climate change. Instituto de Políticas y Bienes Públicos (IPP) CSIC, Working Paper. 2020-03

Available at: digital.csic.es

Pledge and implement bargaining in the Paris Agreement on climate change

Alejandro Caparrós*

April 30, 2020

Abstract: This paper analyzes a multilateral bargain game motivated by the Paris Agreement on climate change. Countries submit pledges, which can be revoked although this implies reputational costs. Pledges, which do not need to be accepted by other countries, detail intended abatement efforts and can be conditional or unconditional, according to whether they depend on transfers. As the process is repeated, incomplete long-term provisions are also considered. The analysis shows the conditions under which, despite its weakness, the Paris Agreement can bring the world to the first best, or at least closer. It also details how to improve the current agreement.

JEL classifications: C73, C78, H41, Q54.
Key words: Bargaining Theory, Stochastic and Dynamic Games, Global Warming.

*Institute for Public Goods and Policies (IPP), Spanish National Research Council (CSIC), Spain (email: alejandro.caparros@csic.es). I thank Scott Barrett, Michael Finus, Bart Harstad, Jean-Christophe Pereau, Santiago Rubio, Sareh Vosooghi, seminar audiences, and participants to the WCERE2018, FRA2018 and AERNA2019 conferences/workshops for helpful comments and discussions. Financial support was provided by the Spanish Ministry of Science and Innovation (project TrenGood ECO2017-84461-R).

1 Introduction

As stated by William Nordhaus in his Nobel Prize address, "global warming is the most significant of all environmental externalities" and "the ultimate challenge for economics" (Nordhaus, 2019). As he also pointed out, "it is only by designing, implementing, and enforcing cooperative multinational policies that nations can ensure effective climate-change policies". In a recent effort to do so, most countries of the world signed the Paris Agreement in 2015, and ratified it soon thereafter. According to this agreement, countries aim at "holding the increase in the global average temperature to well below 2°C above pre-industrial levels and [..] pursue efforts to limit the temperature increase to 1.5°C above pre-industrial levels" (UNFCCC, 2015). The Paris Agreement has no short-term commitments, but rather urges countries to submit their voluntary commitments (pledges), their Nationally Determined Contributions (NDCs). NDCs are not legally binding and are neither subject to the acceptance by other countries, nor to the unanimity rule needed to adopt international agreements, such as the Paris Agreement itself.

Rogelj *et al.* (2016) published an analysis determining the contribution of the submitted NDCs[1] to the 2°C target. The analysis shows that current climate policies cover about 25% of the required effort and that current NDCs would add another 25%, so that roughly 50% of the required effort is missing[2]. In addition, about one third of the NDCs are conditional, requiring, for example, the provision of international funds in the case of many developing countries; and an additional 45% of the NDCs came with both conditional and unconditional components. Thus, only about 20% of NDCs are purely unconditional (i.e. pledging to abate emissions by a certain amount irrespective of the behavior of others). In other words, even if countries meet their (unconditional) NDCs there is a need for substantial additional effort to reach the 2 °C target (let alone the 1.5 °C target).

Although initial attention focused mainly on the role of the NDCs, the Paris Agreement has a second key feature: the mechanism to foster cooperation provided by articles 6 and 9, which may become the most important articles in the long term (Mehling *et al.*, 2018; Schneider *et al.*, 2019). Article 6 allows countries to cooperate voluntarily to increase the ambition of their abatement targets.[3] Cooperation between countries can significantly reduce

[1] Rogelj et al. (2016) analyzed Intended NDCs, submitted just before the Paris summit. NDCs finally submitted were similar to the Intended NDCs in most cases.

[2] Along the same line, see the report of the Intergovernmental Panel on Climate Change (IPCC) on Global Warming of 1.5° (Masson-Delmotte *et al.*, 2018).

[3] See Mehling *et al.* (2018) for a discussion on how what they call 'mitigation output transfers' could operate in the Paris Agreement under article 6.2 , including the role of outsiders to the Paris Agreement. On the role of outsiders, notably the US as it has announced its intention to withdraw from the Agreement, see Jotzko *et al.* (2018).

costs and allow for more ambitious goals. Agreements based on this article can be legally binding and will probably have a shorter time horizon. Article 9 is also relevant as it considers monetary transfers from developed countries to assist developing countries in their efforts to mitigate and adapt to climate change. Developed countries have agreed to mobilize a substantial amount of money: "prior to 2025 the Conference of the Parties [...] shall set a new collective quantified goal from a floor of USD 100 billion a year".

The pledges system instituted in the Paris Agreement is usually referred to as the 'pledge-and-review' mechanism. This terminology was used in the early nineties when debating about the appropriateness of moving from the 'pledge-and-review' system introduced in the United Nation Framework Convention on Climate Change (UNFCCC) to the burden-sharing approach instituted in the Kyoto Protocol (a protocol to the UNFCCC). However, the system introduced in the Paris Agreement (another development of the UNFCCC) goes beyond a mere pledge-and-review mechanism. In addition to articles 6 and 9 mentioned above, at the Conference of the Parties (COP) in Katowice in 2018 national negotiators were charged with developing the so-called 'rulebook' to fully shape and specify the process set out by the Paris Agreement. The so-called 'Katowice package' is so detailed that some have argued that is has brought the Paris Agreement framework closer to a Kyoto-like framework. I therefore refer to the mechanism created by the Paris Agreement as 'pledge-and-implement', in the sense that, after the submission of the pledges, additional negotiations should be expected during its implementation, mainly as developments of articles 6 and 9 (the details of these articles are still under discussion and were the main topic during the COP in Madrid, in 2019).

To illustrate this, let us focus on a developing country that, in its conditional pledge, requested a transfer in exchange for an additional reduction in its greenhouse gas emissions beyond what it had pledged to do in its unconditional pledge. As detailed below, in my model the unconditional pledge corresponds to the Nash equilibrium (NE) without any interaction, i.e. what the country would do based on its own interest and taking the behavior of other countries as given, and the conditional pledge will be interpreted as a request for a share of the potential additional surplus created (e.g. through a transfer, see footnote 3). Then, once all the pledges are received, and developed countries have pledged to contribute a given amount through art 9, additional negotiations take place to determine the transfers finally offered to the developing country at the implementation phase (mainly through article 6). If the transfer resulting from this negotiation equals the one requested, fine. However, if the transfer coming out of the negotiation is lower, the government of our developing country needs to come back to its constituency and explain that they are making the additional effort for less, i.e. that they obtained a smaller share. This has costs, which can be political or reputational (be it national or international reputation). Hence, in my model, obtaining a final share which is smaller than the one requested in the conditional pledge has a cost.

In addition to this 'pledge-and-implement' mechanism, the Paris Agreement contains long-term provisions. The rulebook and the Katowice package mentioned above contain provisions that will govern the pledge-and-implement mechanism in the long-term (several of these long-term provisions, mainly procedural ones, are binding under international law). In general, all the provisions included in the Paris Agreement under the so-called 'ambition cycle', designed to increase the ambition of the agreement, are related to the long-term provision considered here. Furthermore, according to article 4 of the Paris Agreement[4] "all Parties should ... communicate long-term low greenhouse gas emission development strategies".

To include these long-term provisions in my model, in the second part of the paper I assume that countries can sign, or not, a long-term provision that sets out some rules that govern future short-term interactions. I consider that this long-term provision is incomplete in a double sense[5]. First, countries can agree on abatement levels and associated transfers, but they cannot negotiate and sign agreements over investments (no provision in the Paris Agreement indicates how to reach the goals). To avoid confusion, I call agreements which are incomplete in this sense 'non-comprehensive'. Second, the long-term agreement on abatement levels and associated transfers is complete if it is contingent on a stochastic variable that summarizes climate uncertainty, and is incomplete when it cannot depend upon the stochastic component. Real-world treaties are unlikely to include obligations that depend on the realization of a stochastic variable. Even over an issue such as climate change, where uncertainty is a key ingredient of the problem, countries are unlikely to agree on transfers and abatement efforts that depend on the severity of climate change. Moreover, the problem is so complex and multidimensional that it is probably not possible to write a meaningful agreement as a function of the state of nature. I call the long-term provisions which are incomplete in this sense 'non-contingent'.

The analysis shows that the pledge-and-implement mechanism can implement the first-best in the short-term, and that the surplus is ultimately shared according to the relative importance of the cost-of-revoking the pledges (countries with relatively larger costs-of-revoking obtain a larger share). However, without a long-term provision there is underinvestment. The reason for this is the standard hold-up problem, as countries underinvest to improve their bargaining power during the pledge and implement phases. However, if it were pos-

[4] The COP, by its decision 1/CP 21, invited Parties to communicate, by 2020, mid-century, long-term low greenhouse gas emission development strategies in accordance with article 4 of the Paris Agreement. See https://unfccc.int/process/the-paris-agreement/long-term-strategies

[5] Barrett (2002) uses the term 'incomplete' in a third sense, referring to agreements signed only by a subset of all the countries.

sible to write a long-term agreement contingent on the state of nature the first-best would be attained. This holds even if the long-term provision can be abandoned at any time if not profitable for all parts, as I will assume below (countries can always abandon treaties with no penalties, as Canada abandoned the Kyoto Protocol just before the end of the first commitment period). Thus, if the long-term provision could be made contingent on the state of nature it would be able to solve the hold-up problem, despite its weakness. However, as it is probably not possible to write a long-term provision that is truly contingent on the state of nature, I also analyze the role that a non-comprehensive and non-contingent agreement can have. Not surprisingly the results show that the first-best can no longer be attained. Nevertheless, even this form of long-term provision brings the world closer to the first-best. Furthermore, the analysis shows that the problem lays not so much in the non-comprehensiveness of the long-term provision but more on the fact that it is non-contingent. Thus, the more dependent the terms of the long-term provision is on the state of nature finally realized (e.g. on the severity of the climate change observed), the closer the agreement brings us to the first best.

A large body of literature on game-theoretic analyses of climate change agreements has emerged over the last three decades. A recent survey and a selection of the most relevant papers can be found in Finus and Caparrós (2015). Although some papers have focussed on the negotiation process (Caparrós *et al.* 2004; Caparrós 2016), most of this literature abstracts from the negotiation process and focuses instead on the concept of internal and external stability of a coalition (agreement), in the sense that no player wants to leave or join the coalition (Barrett 1994). In any case, for the discussion below, the key point is that none of these papers is appropriate to model the process described above: a long-term provision followed by a series of interactions governed by the pledge-and-implement mechanism. In all these papers the focus was on a long-term burden sharing agreement, such as the Kyoto Protocol.

In terms of motivation, the contemporaneous paper[6] by Harstad (2020) is that which is most related, as it also attempts to model the pledges in the Paris Agreement. The key difference is that Harstad assumes that pledges need to be accepted by other countries, and that countries need to submit a new pledge if at least one country rejects any of the pledges submitted (he does not differentiate between conditional and unconditional pledges). Interesting as this option would be, a repetition of the submission of pledges if they are not accepted by all the countries that contribute positively is not a feature of the Paris Agreement. It is true

[6] See also Harstad (2019), where he embeds the bargaining model developed in Harstad (2020) into a dynamic climate game; and the experimental analysis of the pledges in the Paris Agreement in Barrett and Dannenberg (2016).

that new agreements by the COP are usually subject to the unanimity rule (or at least to an absence of an objection). In fact, this was the reason why the Copenhagen Accord was not adopted as a decision of the COP in 2009 (because of the opposition of Sudan, Venezuela, Bolivia and Nicaragua). It is also true that the Copenhagen Accord included pledges, which were therefore subject to the unanimity rule at the time. However, the Paris Agreement has followed a different path. First, the Paris Agreement was adopted unanimously, i.e. without any objection, as a decision of the COP in 2015 (UNFCCC, 2015). Pledges were then submitted by the different countries, from 2016 onwards[7], and these pledges are explicitly excluded from the unanimity rule, as countries merely "communicate" their pledges (NDCs), which are recorded in a public registry[8] by the secretariat. The same holds for future pledges under the Paris Agreement. Hence, the pledges that I am modeling here are not subject to any approval and cannot be objected to.

The bargaining part of my model, the pledge-and-implement mechanism, is related to the bilateral bargaining model with partial commitment introduced by Muthoo (1996), in the context of the Nash Bargaining Solution, and analyzed using Rubinstein's alternating-offers model in Leventoglu and Tarar (2005). Focusing on this part, and apart from the integration in a long-term model, my contribution is to extend the latter model to more than two players. This is not a trivial task in a model with alternating offers like mine (using the NBS and extending Muthoo's approach would also not be trivial, as he proposes a graphical proof).

Precedents to my analysis of the long-term provisions can be found in Harstad (2016), and to a lesser extent in Beccherle and Tirole (2011), Harstad (2012) and Battaglini and Harstad (2016). Although in a different framework, Harstad (2012 and 2016) analyzes incomplete contracts in the first sense discussed above, i.e. in the sense that countries can negotiate and sign agreements on emissions and not on investment (what I have called non-comprehensive). Beccherle and Tirole (2011) focus on the hold-up problem in a two periods model and Battaglini and Harstad (2016) extend the analysis to include the role of participation and duration of the agreements. Harstad (2012) and Battaglini and Harstad (2016) analyze a deterministic framework, while Harstad (2016) includes a stochastic component in a model with a quadratic specification, which is therefore more relevant for my analysis. Although he considers long-term agreements, he does not consider the two-tier procedure described above, one long-term provision followed by a sequence of short-term interactions, nor does he focus on agreements with partial commitment (pledges) as I do here. Other relevant precedents can be found in Dutta and Radner (2004, 2009, 2012) and in Barrett (2002). The

[7] 'Pledges' submitted before the approval of the Paris Agreement were called *Intended* NDC, instead of NDCs, and had no legal status.

[8] Article 4.12: "Nationally determined contributions communicated by Parties shall be recorded in a public registry maintained by the secretariat" (UNFCCC, 2015).

former focus on equilibrium selection in analyses that combine theory and simulations and the latter focuses on the role of alternative assumptions when analyzing the possibility to renegotiate agreements. Building on their previous models, Dutta and Radner (2019) show in a recent paper that a well defined transfer, the Green Climate Fund in their analysis, can bring the world to the first best, using reversion to Business as Usual as a threat, and a fixed order of talking that gives every player Stackelberg-leader features (i.e. 'followers' do not have the possibility of rejecting an offer and making a counteroffer at a later stage). Although they also refer to the Paris Agreement, their focus differs from mine in that they do not focus on the partial commitment strategies introduced by the NDCs (pledges) nor on the interaction of this type of commitment and long-term provisions.

The part devoted to the long-term provision is also closely related to the vast literature on incomplete contracts (Muthoo 1995, Hart and Moore 1988, Hart and Holstrom 1987 or Crawford 1988), and several of the results shown below are well known in that literature (see Segal and Whinston 2010 for a survey). This part of my model is basically a hold-up model, although it does not assume self-investments (see the definition below) as most of these models do (generally as a consequence of the assumption that no trade takes place if there is no agreement). My model, in contrast, has imperfect coalition investments, despite the fact that investments only reduce the cost function of the country investing. The reason for this is that, due to the public good nature of the abatement undertaken in the last stage, when one country increases its investment it improves the disagreement point for the remaining countries.

The rest of the paper is organized as follows: Section 2 presents the multilateral bargaining model with partial commitment, and conditional and unconditional pledges; Section 3 integrates this bargaining model in a multi-stage model where countries invest in long-term assets before entering the bargaining phases; Section 4 discusses the results and Section 5 concludes.

2 Pledge and implement bargaining

There are N countries which can provide abatement efforts to mitigate a 'public bad' problem, such as climate change (i.e., abatement is a public good). There are two phases: a first phase where countries submit their pledges simultaneously (their NDCs) and a second phase where the pledges are (eventually) implemented. In the first phase ('pledge' phase) countries submit simultaneously their conditional and their unconditional pledges (NDCs). Unconditional pledges are independent of the behavior of other countries. Conditional abatement efforts are additional abatement efforts which are conditional on transfers, or on the behavior of other countries. If all countries submit only unconditional pledges, at the second phase each country decides independently its abatement efforts. If at least one country submits a conditional pledge, in the second phase ('implementation' phase) countries negotiate the implementation of the conditional pledges. Pledges can be revoked but at a certain cost, as detailed below. That is, backing off from an initial pledge has costs and countries suffer a greater loss the greater the distance from their initial pledge to the agreement finally implemented. This cost can be seen as reputational (national or international reputation, see the introduction).

Gross per-period payoffs are given by the function

$$v_i(\mathbf{q}) = B_i(Q) - C_i(q_i),\ \forall i \in N \tag{1}$$

where $\mathbf{q} = (q_1, ...q_N)$ shows the abatement effort finally undertaken by each country, B_i summarizes benefits from climate change abatement and C_i shows the cost incurred through the abatement efforts. Benefits B_i depend on aggregate abatement ($Q = \sum_{i \in N} q_i$). Costs C_i depend on the amount of abatement performed by county i. The discount rate is given by $r > 0$, with $\delta \equiv e^{-r\tau}$. Regarding the payoff function mentioned above, I make the following standard assumptions:

Assumption 1.a: *For all i,j* $\in N$: $\frac{\partial B_i}{\partial q_j} > 0$, $\frac{\partial^2 B_i}{\partial q_j^2} \leq 0$, $\frac{\partial C_i}{\partial q_i} > 0$, $\frac{\partial^2 C_i}{\partial q_i^2} > 0$, $\lim_{Q \to 0} \frac{\partial B_i}{\partial q_i} > \lim_{q \to 0} \frac{\partial C_i}{\partial q_i} > 0$.

To provide a benchmark, let use denote the Nash equilibrium of the game without interactions, where each country solves $\arg\max_{q_i} v_i(q_i, \mathbf{q}_{-i})$, as $\mathbf{q}^A = (q_1^A, ...q_N^A)$. This is the equilibrium without any pledges or any negotiations. However, given assumption 1.a this implies a positive level of abatement for all countries.

As 'unconditional' pledges (NDCs) are independent of the behavior of other countries, each country simply submits its own pledged abatement effort q_i^U (recall that NDCs do not need to be accepted by other countries). The vector of all unconditional pledges is noted: $\mathbf{q}^U = (q_1^U, ...q_N^U)$. A 'conditional' pledge (NDC) by country i is a pair formed by (i) an abatement vector, noted $\mathbf{q}^C = (q_1^C, ...q_N^C)$, where $q_i^C \geq q_i^A\ \forall\ i$ and at least for one country $q_j^C > q_j^A$, and (ii) a requested share for country i of the additional surplus created equal to α_i. That is, for at least one country the abatement proposed in the conditional pledges is above the NE without interaction, and the country submitting the conditional pledge requests a share of the additional surplus created. For example, a conditional pledge from country i can request a transfer in exchange for an additional abatement, as proposed by several developing countries (see the introduction).

Before continuing, let us note first that the abatement efforts included in the unconditional pledges are identical to those associated with the Nash equilibrium of the game without any negotiations, and that, on the aggregate, these abatement efforts fall short of the short-term first-best abatement effort, noted $\mathbf{q}^* = (q_1^*, ...q_N^*)$. For future reference, the following Lemma highlights this straightforward result:

Lemma 1 *Abatement efforts included in the unconditional NDC are identical to those corresponding to the Nash Equilibrium of the game without interaction, i.e.* $\mathbf{q}^U = \mathbf{q}^A$. *These abatement efforts are smaller than the short-term first-best efforts at the aggregate level, i.e.* $\sum_{i \in N} q_i^U < \sum_{i \in N} q_i^*$.

Proof. Each country solves $\arg\max_{q_i} v_i(\mathbf{q}, I_i, \theta)$ taking $\mathbf{q}_{-i}$ as given. This yields the standard Nash Equilibrium:

$$\frac{\partial B_i}{\partial q_i} - \frac{\partial C_i}{\partial q_i} = 0 \Longrightarrow q_i^U = q_i^A \quad \forall i \in N \tag{2}$$

Short-term first-best abatement efforts are given by

$$\sum_{i \in N} \frac{\partial B_i}{\partial q_i} - \frac{\partial C_i}{\partial q_i} = 0 \Longrightarrow q_i^* \quad \forall i \in N \tag{3}$$

and it is straightforward to show that they are larger at the aggregate level. ■

Because conditional pledges imply, by definition, that at least one country will abate more if implemented, conditional pledges create an additional surplus. Let us denote this additional surplus by

$$\Delta(\mathbf{q}^A, \mathbf{q}^C) = \sum_{i \in N} \left[v_i(\mathbf{q}^C) - v_i(\mathbf{q}^A) \right].$$

As noted above, pledges can be revoked but at a certain cost. To avoid unnecessary complications, I assume that unconditional pledges are never revoked (as Lemma (1) has shown, unconditional pledges just write down the efforts associated to the Nash equilibrium without any interaction). For conditional pledges, I assume the following functional form for the cost-of-revoking function:

$$K(\alpha_i, \varphi_i) = \begin{cases} 0 \text{ if } \alpha_i \leq \varphi_i \\ k_i(\alpha_i - \varphi_i) \text{ if } \alpha_i > \varphi_i \end{cases} \tag{4}$$

where $k_i > 0$ and φ_i is the share finally implemented. That is, countries suffer more the larger the distance between the share requested in their conditional NDC and the share finally obtained. To keep the model as simple as possible, I assume a linear relationship. If countries are unable to agree at the implementation phase, the disagreement point is given by the unconditional pledges.

At the implementation phase, countries either implement their unconditional abatement efforts (Nash equilibrium without interactions) or sign agreements based on articles 6 and 9 to implement the additional efforts included in the conditional pledges. These agreements are negotiated following a standard offer and counter-offer procedure, but taking into account that modifying the position set out in the pledges has a cost. To simplify, I assume a cyclical protocol where the order in which countries talk is known (the order in which countries talk captures the relative importance of different countries in the international arena; however, section 4.1.1 relaxes this assumption). I also assume that countries talk in the order given by their ordinality (as I can rename countries at will this implies no loss of generality).

Summing up, countries submit their pledges simultaneously and, if all countries submit only unconditional pledges, or if (all or some) countries submit conditional pledges but there is no agreement at the implementation phase, the net payoff for country i is given by:

$$\Pi_i(\alpha_i, \varphi_i, \mathbf{q}^A) = v_i(\mathbf{q}^A)$$

If (all or some) countries submit conditional pledges and there is an agreement at the implementation phase, the net payoff for country i is given by:

$$\Pi_i(\alpha_i, \varphi_i, \mathbf{q}^A, \mathbf{q}^C) = v_i(\mathbf{q}^A) + \varphi_i \Delta - K(\alpha_i, \varphi_i) \ \forall\ i \in N \tag{5}$$

Restricting attention to Stationary Subgame Perfect Equilibria (SSPE) to simplify the analysis, the following proposition describes the equilibrium NDC pledges and the associated payoffs:

Proposition 1 *Each country submits an unconditional NDC where q_i^U is defined by the Nash equilibrium of the game without interactions, as defined in equation (1), i.e. $q_i^U = q_i^A$; and a conditional NDC consisting of a set of abatement efforts $\mathbf{q}^C = \mathbf{q}^*$ and associated transfers that imply a share for country i of the additional surplus created, $\Delta(\mathbf{q}^A, \mathbf{q}^C)$, equal to*

$$\alpha_i^* = \frac{\delta^{i-1}(1+k_i)}{\sum_{i=1}^{N} \delta^{i-1}(1+k_i)} \forall i \in N. \tag{6}$$

In the unique equilibrium, the short-term first-best abatement efforts, $\mathbf{q}^C = \mathbf{q}^*$, are implemented and the shares finally implemented are $\varphi_i^* = \alpha_i^*\ \forall i \in N$, with $\sum_{i=1}^{N} \varphi_i^* = 1$. Thus, no country needs to back-up from its initial conditional NDC and net payoffs are given by:

$$\Pi_i = v_i(\mathbf{q}^A) + \varphi_i^* \Delta(\mathbf{q}^A, \mathbf{q}^*) \ \forall\ i \in N,$$

as $K(\alpha_i^*, \varphi_i^*) = 0$ for $\alpha_i^* = \varphi_i^*$.

Proof: Appendix A.1

As discussed above, countries agree on the abatement level that maximizes the additional surplus to be shared, and each country obtains a share φ_i of the additional surplus created, plus its payoff in the situation where all countries follow their unconditional NDC (the Nash equilibrium without interactions). Although the detailed proof can be found in Appendix A.1, the intuition that explains why countries implement the first-best abatement effort is as follows. Any set of abatement efforts that falls short of the socially optimal abatement effort cannot be an equilibrium, because countries are interested in increasing the ambition to avoid leaving unexhausted surplus. In equilibrium, conditional pledges can also not include a set of transfers that implies that the sum of the shares requested of the surplus created exceed one. The reason for this is that at least one country would be interested in requesting a smaller share, to avoid revocation costs. A set of transfers that implies that the sum of the shares requested is less than one can also not be an equilibrium, as at least one country would be interested in requesting a larger share. Thus, abatement efforts and requested shares lay in equilibrium on the Pareto frontier. What Appendix A.1 shows is that only one point on the Pareto frontier can be sustained as an equilibrium. In that point, the additional surplus obtained by moving from the disagreement point (where all countries implement their unconditional NDC) to the situation where all implement their conditional NDC, is shared proportionally to the cost-of-revoking function of each country. Countries that suffer higher costs from modifying their initial pledges obtain a larger share of the surplus created. Nevertheless, these costs-of-revoking never actually occur in equilibrium, as in our world countries submit 'perfect' conditional NDC pledges.

As already mentioned, the share obtained from the additional surplus depends on the relative cost to the country of modifying its position. If all have the same cost-of-revoking function, the 'split the difference' rule holds and $\varphi_i = \varphi = 1/N \ \forall i$. If the cost for one country of changing its conditional contribution tends to infinity, while the costs of the remaining countries is finite, the uncompromising country obtains all the surplus. Note also that, the total abatement achieved with the 'pledge and implement' procedure is first best, and identical to the one obtained using the asymmetric Nash bargaining solution. However, the distribution of the surplus generated is not identical. The result that the pledge-and-implement mechanism leads essentially to the same results as applying directly the Nash Bargaining Solution could be seen as implying that the Paris Agreement is essentially a burden sharing mechanism. However, this is only true if all countries use conditional NDC and articles 6 and 9 are fully developed, allowing for a meaningful implementation phase. Section 4.3 discusses the extent to which this is the case in reality.

3 Long term provisions and investment

The previous section has shown that the 'pledge and implement' mechanism can bring abatement levels to the first best in the short term. In this section I analyze the role of long-term provisions in the Paris Agreement. Long term provisions are relevant because through long-term investments countries can reduce the costs of short-term abatement efforts. In addition, I incorporate uncertainty in the model, a key feature of the climate change problem. As discussed in the introduction, the Paris Agreement and its recent developments contain several long-term provisions.

The extended model has four stages, as shown in Figure 1. At stage 1, countries decide to sign, or not, a long term 'non-comprehensive' provision. The long-term provision is non-comprehensive because it cannot specify investment levels. That is, the long-term provision only specifies the additional abatement that countries will perform at stage 4, and the associated transfers, as detailed below. Note that this is consistent with the fact that investments are not covered in any form in the Paris Agreement. In addition, this long-term provision can be 'contingent', if it is written as a function of the state of nature described below, or 'non-contingent', if it is not contingent on the state of nature.

[Figure 1 about here]

Whether or not a long-term provision is included in the Paris Agreement (stage 1), countries decide at stage 2 their investment levels, denoted by $\mathbf{I} = (I_1, ...I_N)$. Investing I_i units has a cost equal to I_i for country i. This investment is private and only benefits directly the country investing, by reducing its abatement cost in the subsequent stages.

At stage 3, the state of nature θ is randomly realized, out of the finite set of possible states of nature Θ. The probability of each state of nature is known. After the value of θ is observed at stage 3, it becomes common knowledge.

Stage 4 has an infinite number of identical periods. Each one of these periods can be divided into the two phases described in the previous section. The only relevant modification to the analysis shown above is that gross payoffs[9] per period are now given by the function

$$v_i(\mathbf{q}, I_i, \theta) = B_i(Q, \theta) - C_i(q_i, I_i),\ \forall i \in N.$$

This function is similar to the function introduced in equation (1) except that now benefits B_i depend on aggregate abatement ($Q = \sum_{i \in N} q_i$) and on the realization of a random variable θ, and costs C_i depend on the amount of abatement performed by county i at stage 4 and on

[9]Note that none of the results shown below depend on the separable nature of benefits and costs, as I could work directly with the function v. Nevertheless, I use functions B_i and C_i to facilitate the interpretation and comparability with the rest of the literature on climate agreements.

the investment made by county i at stage 1 (investment stage). In addition to Assumption 1.a, I make the following assumptions:

Assumption 1.b: *For all* $i \in N$: $\frac{\partial C_i}{\partial I_i} < 0, \frac{\partial^2 C_i}{\partial I_i^2} > 0, \frac{\partial^2 C_i}{\partial q_i \partial I} < 0, \frac{\partial B_i}{\partial \theta} > 0.$

At the Nash equilibrium without any interactions of this extended model all countries solve $\arg\max_{q_i} v_i(q_i, q_{-i}, I_i, \theta)$ in each period of stage 4, once $\mathbf{I}$ and θ are known. This yields

$$\frac{\partial B_i}{\partial q_i} - \frac{\partial C_i}{\partial q_i} = 0 \Longrightarrow \; q_i^A = q_i^A(I_i, \theta), \; \forall i \in N. \tag{7}$$

At stage 3 there are no strategic decisions and at stage 2 country i solves $\arg\max_{I_i} E_\theta \left[\frac{v_i(\mathbf{q}^A, I_i, \theta)}{1-\delta}\right] - I_i$, yielding

$$E_\theta \left[- \left. \frac{\partial C_i}{\partial I_i} \right|_{q_i^A(I_i, \theta)} \right] = 1 - \delta \; \Longrightarrow I_i^A, \quad \forall i \in N. \tag{8}$$

Because all periods at stage 4 are identical and $\mathbf{I}$ and θ are already known at the beginning of stage 4, the analysis of the pledge-and-implement process shown in section 2 is valid for any of the periods in stage 4 if no long term provision is signed at stage 1 (substituting $v_i(\mathbf{q})$ by $v_i(\mathbf{q}, I_i, \theta)$). That is, if no long-term provision is signed, countries submit their NDC pledges freely and the pledge-and-implement process is repeated in each period. If there is perpetual disagreement at the implementation phase in one period, countries follow the Nash equilibrium without interaction from then onwards (their unconditional pledges). If a long-term provision was signed, as detailed below, countries follow its terms if they are profitable for all parts, taking into account the value of θ observed at stage 3 (otherwise, countries are free to send their NDC pledges freely again).

Before continuing, I formally define self-investments and cooperative investments for future reference. As already mentioned in the introduction, although investment is a private good in the sense that it only benefits directly the country that undertakes the investment, using the terminology commonly used in the literature on incomplete contracts (Segal and Whinston, 2010) it is a cooperative investment, in the sense that it affects the disagreement payoff of other countries.

Definition 1 *Country i's investment is a 'self-investment' if country j's disagreement payoff, $v_j(\mathbf{q}^A, I_j, \theta)$, is independent of I_i for all $j \neq i$, at all I_{-i} and θ. Country i's investment is 'cooperative' if country j's disagreement payoff is non decreasing in I_i for all $j \neq i$, at all I_{-i} and θ.*

3.1 First best

To provide a benchmark, let us analyze the first-best outcome for this extended model. As all short term interactions are identical, the central planner has the following objective function:

$$f(\mathbf{q}, I_i, \theta) = \arg\max_{\mathbf{I},\mathbf{q}} \left(E_\theta \left[\frac{\sum_{i \in N} v_i(\mathbf{q}, I_i, \theta)}{1-\delta} \right] - \sum_{i \in \mathbf{I}} I_i \right) \tag{9}$$

We can write the following lemma:

Lemma 2 *The first best equilibrium of the extended model in terms of abatement and investment is defined by the following equations:*

$$E_\theta \left[\sum_{i \in N} \frac{\partial B_i}{\partial q_i} - \frac{\partial C_i}{\partial q_i} \right] = 0 \Longrightarrow q_i^* = q_i^*(I_i, \theta) \quad \forall i \in N \tag{10}$$

$$E_\theta \left[- \left. \frac{\partial C_i}{\partial I_i} \right|_{q_i^*(I_i,\theta)} \right] = 1 - \delta \quad \Longrightarrow I_i^* \quad \forall i \in N \tag{11}$$

Proof. The FOC for the abatement efforts are given by:

$$\frac{\partial f}{\partial q_i} = E_\theta \left[\frac{1}{(1-\delta)} \left(\sum_{i \in N} \frac{\partial B_i}{\partial q_i} - \frac{\partial C_i}{\partial q_i} \right) \right] = 0 \quad \Longrightarrow \quad q_i^* = q_i^*(I_i, \theta) \quad \forall i \in N, \tag{12}$$

yielding equation (10). Define the maximum value function:

$$F(\mathbf{q}, \mathbf{I}, \theta) = f(\mathbf{q}^*(\mathbf{I}, \theta), \mathbf{I}, \theta) = f(q_1^*(I_1, \theta), ..., q_N^*(I_N, \theta), I_1, ..., I_N, \theta)$$

where $q_i^*(I_i,\theta)$ are the optimal levels of abatement stage 4. From the envelope theorem, using (12):

$$\frac{\partial F(\mathbf{q}, \mathbf{I}, \theta)}{\partial I_1} = \frac{\partial f}{\partial q_1} \frac{\partial q_1}{\partial I_1} + ... + \frac{\partial f}{\partial q_N} \frac{\partial q_N}{\partial I_1} + \frac{\partial f}{\partial I_1} = \frac{\partial f}{\partial I_1}$$

and the FOC in terms of investment is

$$E_\theta \left[\frac{\sum_{i \in N} \frac{\partial v_i(q_i^*(I_i,\theta), I_i, \theta)}{\partial I_i}}{1-\delta} \right] - 1 = E_\theta \left[\frac{- \left. \frac{\partial C_i}{\partial I_i} \right|_{q_i^*(I_i,\theta)}}{1-\delta} \right] - 1 = 0 \quad \forall i \in N,$$

or equation (11). ■

The first order condition in (10) is identical to (3), just taking into account the role of I_i and θ. It is the standard condition defining optimal abatement, but taking into account the role of the investment made and the fact that abatement benefits have a random component. Condition (11) yields the optimal investment level. The central planner takes into account the impact of the investment of country i on the infinite stream of future costs of abatement. However, as one only needs to take into account direct effects, due to the envelope theorem, and investments are a private good and only benefit directly the country that undertakes the investment, the FOC in terms of investment are independent from each other. Thus, each FOC yields the optimal investment for one country.

3.2 The investment stage without any long-term provision (the 'basic Paris Agreement')

Consider now the situation where there is no long-term provision determining the pledges that countries should submit in each period (this can be seen as a 'basic Paris Agreement', without the long-term provisions discussed in the introduction). To solve the game by backward induction, let us start with stage 4. In this case, in each period of stage 4 the situation is akin to the bargaining problem with two phases analyzed in section 2. However, the surplus created in each period now reads:

$$\Delta(\mathbf{q}^A, \mathbf{q}^C, \mathbf{I}, \theta) = \sum_{i \in N} \left[v_i(\mathbf{q}^C, I_i, \theta) - v_i(\mathbf{q}^A, I_i, \theta) \right].$$

That is, in addition to $\mathbf{q}^A$ and $\mathbf{q}^C$, the share now also depends on $\mathbf{I}$ and θ. Nevertheless, because both are known at the beginning of stage 4, the analysis shown in section 2 holds without any significant modification.

From Proposition 1 we know that at each interaction at stage 4 countries agree on the abatement level that maximizes the additional surplus to be shared. Taking into account the role of I_i and θ this implies that $q_i^C(I_i, \theta) = q_i^S(I_i, \theta) = q_i^*(I_i, \theta)\ \forall i \in N$ is given by (10). At stage 3 there are no strategic decisions (only θ is revealed at this stage). At stage 2 (investment stage), the expected payoff for each country i is:

$$E_\theta \left[\frac{v_i(\mathbf{q}^A, I_i, \theta) + \varphi_i \sum_{j \in N} \left[v_j(\mathbf{q}^*, I_j, \theta) - v_j(\mathbf{q}^A, I_j, \theta) \right]}{1 - \delta} \right] - I_i \tag{13}$$

i.e., its disagreement payoff in each period plus a share of the additional surplus created through the pledge-and-implement bargaining process in each period, minus the cost associated with the investment. The following proposition summarizes the results concerning the investment level (stage 2) for the 'basic Paris Agreement' (S in superscript denotes equilibrium values without any long-term provision).

Proposition 2 *Countries do not implement the first best through a series of short term interactions (pledge-and-implement) without a long-term provision, as investment,* I_i^S*, is* *defined, $\forall i \in N$ and $j \neq i$,by*

$$\underbrace{\underbrace{\varphi_i E_\theta \left[-\left.\frac{\partial C_i}{\partial I_i}\right|_{q_i^*(I_i,\theta)} \right]}_{\textit{Conditional NDC}} + \underbrace{(1-\varphi_i) E_\theta \left[-\left.\frac{\partial C_i}{\partial I_i}\right|_{q_i^A(I_i,\theta)} \right]}_{\textit{Unconditional NDC}}}_{\textit{Self-investment terms}} \tag{14}$$

$$\underbrace{-\varphi_i \sum_{j \in N/i} E_\theta \left[\left.\frac{\partial B_j}{\partial q_i}\right|_{q_i^A(I_i,\theta)} \frac{\partial q_i^A(I_i,\theta)}{\partial I_i} \right]}_{\textit{Cooperative investment term}}$$

$$= \quad 1-\delta$$

Thus, there is underinvestment as compared to the first best and as a consequence abatement is also below the optimal level. Overall underinvestment is more severe the larger the number of countries, and the larger the benefit obtained by other countries from the increased abatement in case of disagreement induced by larger investments. For a particular country, a larger share of the surplus generated by the short-term agreement, φ_i, reduces underinvestment if increasing investment has a stronger impact on payoffs in case of agreement than in case of disagreement.

Proof. From (13) we know that the NE at stage 2 is given by:

$$E_\theta \left[\frac{\partial v_i(\mathbf{q}^A, I_i, \theta)}{\partial I_i} \right] + \varphi_i \sum_{j \in N} E_\theta \left[\frac{v_j(\mathbf{q}^*, I_j, \theta)}{\partial I_i} - \frac{\partial v_j(\mathbf{q}^A, I_j, \theta)}{\partial I_i} \right] = 1 - \delta.$$

Applying the envelope theorem, taking into account that $\frac{\partial v_j(\mathbf{q}^A, I)}{\partial q_i^A} \neq 0$, we obtain (14). As $-\left.\frac{\partial C_i}{\partial I_i}\right|_{q_i^*(I_i,\theta)} > -\left.\frac{\partial C_i}{\partial I_i}\right|_{q_i^A(I_i,\theta)}$ the first two terms on the LHS of (14) imply an underinvestment, compared to the first-best investment given by (11). Given Assumption 1.b, from (7) and (8) we know that $\frac{\partial q_i^A(I_i,\theta)}{\partial I_i} > 0$. Hence, given Assumption 1.a, the last term on the LHS of (14) reduces the equilibrium investment further. Overall underinvestment is more severe the larger the number of countries because the last term on the LHS of (14) becomes larger, for a given φ_i. Larger φ_i reduces underinvestment if

$$E_\theta \left[-\left.\frac{\partial C_i}{\partial I_i}\right|_{q_i^*(I_i,\theta)} \right] > E_\theta \left[-\left.\frac{\partial C_i}{\partial I_i}\right|_{q_i^A(I_i,\theta)} + \sum_{j \in N/i} \left.\frac{\partial B_j}{\partial q_i}\right|_{q_i^A(I_i,\theta)} \frac{\partial q_i^A(I_i,\theta)}{\partial I_i} \right],$$

yielding the last statement in the proposition. ■

As the proposition shows, for a given country, increasing its investment at stage 2 has three effects. First, it decreases its costs in the event of an agreement at stage 4. This is the first term on the LHS of (14). As highlighted in equation (14), this term is related to the conditional NDC, as the derivative of the cost function is evaluated at the abatement level included in the conditional NDC. Given Assumption 1.b, the smaller the share of the surplus created by the agreement that the country appropriates, φ_i, the smaller the amount invested would be. Note that if investment had no impact on the disagreement point only this effect would be relevant, as the second and third terms on the LHS of (14) would vanish. In fact, this is the case in the standard model analyzed in Hart and Moore (1988) or Muthoo (1995), where the no-agreement situation implies no-trade between a buyer and seller. It would also be the case if one assumed, to simplify, that there is no abatement in case of disagreement. Underinvestment would be explained exclusively by the share φ_i that the country receives from the additional surplus created, as the individual country only cares about the share of the additional surplus created through the investment that it appropriates. The second effect is that investment reduces the costs for the country in the event of a disagreement at stage 4. This is reflected in the second term on the LHS in (14), where the derivative of the cost function of this term is evaluated at the abatement level included in the unconditional NDC. Given Assumption 1.b this term favours larger investment. Intuitively, the reason for this is that there is now a second incentive to invest, to reduce the costs in the event of a disagreement. However, the larger the share φ_i obtained by country i the less important this term becomes. These two effects (terms) appear if investment is a self-investment and if it is a cooperative investment (see Definition 1), as they refer to the impact that investment has on the agreement and disagreement payoffs of the country itself. Note also that, for the reasons discussed in the proof of Proposition 2, the net effect of these two effects is an underinvestment, compared to the first-best level. The third effect, reflected in the last term on the LHS of (14), appears only because investment is cooperative in this model. The investment made by country i at stage 2 impacts the disagreement point of all the other countries. The reason for this is that, although investment only impacts the cost function of the investing country, the investment implies more abatement by country i at stage 4 and this improves the disagreement point of all the other countries. This term becomes more important the larger the share of the surplus obtained by the country, and reduces the investment level of country i (for the reasons discussed in the proof of Proposition 2). Intuitively, improving the disagreement point of the other countries reduces the bargaining power of the country investing, and hence its incentive to invest.

Figure 2 illustrates equations (11) and (14), using the functional forms and the parameter values used in Example 1 (see below). The thick solid horizontal line is the RHS of (11) and (14), i.e. $1-\delta$. The dashed line is the LHS of (11). The first-best investment level, I^*, is given by the intersection of these two lines. The line with dots is the first term on the LHS of (14). The line with circles includes the first two terms on the LHS of (14). Finally, the thin solid line includes all the terms on the LHS of (14). The intersection of the thin solid line with the one representing $1-\delta$ yields the investment in the 'basic Paris Agreement'.

[Figure 2 about here]

The following example illustrates the model using particular functions that are similar to those usually used in the literature on climate agreements (Barrett, 1994).

Example 1 *Consider $N \geq 3$ identical countries, implying $\varphi_i = \varphi = 1/N\ \forall i$. Payoffs are $B(Q,\theta) = b\theta Q$ and $C(q,I) = \frac{c}{2I}\left(q^2+F\right)$, where b, c and F are non-negative parameters, with $E_\theta\left[N^2b^2\theta^2\right] < 2(1-\delta)c$. The states of nature in Φ are $\{0.5, 1, 1.5\}$, with equal probability. Linear benefits are used frequently in this literature, even in papers where they are assumed to capture the infinite stream of future benefits. In the context analyzed here, $B(Q,\theta)$ only captures benefits in one period, and a linear specification is therefore more appropriate. The abatement cost function is also similar to the one frequently used in this literature (identical for $I=1$), although with an additional fixed cost F. However, in this model investment at stage 1 can reduce abatement costs (for $I>0$). Dropping subscripts, optimal abatement and investment levels at the first best are*

$$\begin{aligned} q^*(I,\theta) &= \frac{Nb\theta I^*}{c} \\ I^*(\theta) &= \left(\frac{Fc^2}{2c\left(1-\delta\right)-N^2b^2\theta^2}\right)^{\frac{1}{2}}. \end{aligned}$$

For the case of a series of pledge-and-implement interactions, but with a NE at the investment stage with no long-term provision ('basic Paris Agreement'), we have:

$$\begin{aligned} q^S(I,\theta) &= \frac{Nb\theta I^S}{c} \\ I^S(\theta) &= \left(\frac{Fc^2}{2c\left(1-\delta\right)+b^2\theta^2\left(N-3+1/N\right)}\right)^{\frac{1}{2}}. \end{aligned}$$

As predicted by the general analysis, we have underinvestment: $I^S < I^*$. To see that $I^S < I^*$, subtract the denominator from the expression for I^S from the denominator of the expression for I^*. This yields $\frac{1}{N}b^2\theta^2\left(N-1\right)\left(N^2+2N-1\right)$, which is positive. Note also that $I^S > 0$ for $N \geq 3$ and that we also have $q^S < q^*$. As already noted, Figure 2 is drawn using the functional forms in this example, with $(N,b,\theta,c,F,\delta) = (3,1,1,50,3,0.9)$. For these parameter values, in terms of investment we have $I^*(1) = 87$ and $I^S(1) = 31$, and in terms of abatement $q^*(I^*,1) = 5.1962$ and $q^S(I^S,1) = 1.8766$. That is, although the expressions for abatement are similar in both cases, underinvestment implies a significantly lower level of abatement if there is no long-term provision. If there is no long-term provision, abatement *at the disagreement point would be* $q^A(I^S, 1) = 0.6255$ *while abatement at the disagreement point with the first-best investment level would be* $q^A(I^*, 1) = 1.7321$.

3.3 Non-comprehensive but contingent long-term provision

I now assume that countries include a long-term provision in the Paris Agreement, written as a function of θ. This long-term provision, written at stage 1, before θ is known, specifies the transfers that each country will receive and the abatement that they will undertake in each short-term interaction implemented at stage 4. This long-term provision is not binding in the sense that countries can break it. Note that I am here referring to provisions included in the Paris Agreement, which is in general not binding, not to the potential short-term contracts written by developing articles 6 and 9 of the Paris Agreement. If countries want to renegotiate, they simply abandon the long-term provision and negotiate freely following the pledges and implementation phases described in section 2. To simplify the analysis, I assume that writing a new long-term provision after stage 1 is not possible. This can be justified by the effort needed to negotiate a new long-term agreement like the Paris Agreement. Note, however, that this is not really a relevant restriction, as a new long-term provision after investments were made at stage 2 and uncertainty was revealed at stage 3 would not add any strategic component to the sequence of short-term interactions.

Let us fix an arbitrary long-term provision $\langle \tau(\theta) \rangle_{\theta \in \Theta}$ such that for any θ in each short interaction at stage 4:

$$v_i(\mathbf{q}^L, I_i, \theta) + \tau_i(\theta) \geq v_i(\mathbf{q}^A, I_i, \theta) \quad \forall i \in N, \tag{15}$$

where L in superscript stands for contingent long-term agreement. By analogy to the analysis in section 2, we know that abatement in case of agreement is optimal for a given level of investment, as countries have no interest in leaving unexhausted surplus, i.e. $q^L(I_i, \theta) = q^*(I_i, \theta)$. Given my assumptions on v_i there are many long-term provisions that satisfy the restrictions in (15). Let us further assume that the terms of the long-term provision will only be abandoned if it is not profitable for all parts, in the sense that at least one country can credibly threaten to prefer the situation that would prevail without any short-term interactions. Note that the long-term provision would be part of the Paris Agreement, or its developments (such as the Katowice package, see the introduction) and, unlike the pledges, would therefore be subject to the unanimity rule. Abandoning the long-term provision implies no punishment, but it does imply the re-negotiation of abatement efforts. As the outcome of this re-negotiation may, in principle, imply the absence of agreement in future short-term interactions, at least one country should be able to threaten that it prefers the situation without any agreement. That is, if (15) were to fail, the pledge and implement mechanism would go back to the situation without a long-term provision. However, unlike in the situation analyzed in the next section, there are many long-term provisions for which condition (15) would never fail, and I assume that $\langle\tau(\theta)\rangle_{\theta\in\Theta}$ is such a provision.

That is, the contingent long-term provision $\langle\tau(\theta)\rangle_{\theta\in\Theta}$ defined above will never be denounced, as it ensures that (15) holds for any realization of θ. Thus, the expected payoff for country i for any $\mathbf{I}$ at stage 2 is:

$$E_\theta\left[\frac{v_i(\mathbf{q}^*, I_i, \theta) + \tau_i(\theta)}{1-\delta}\right] - I_i \quad \forall i \in N \tag{16}$$

Hence, the NE at the investment stage is given by

$$E_\theta\left[\frac{\partial v_i(\mathbf{q}^*, I_i, \theta)}{\partial I_i}\right] = 1-\delta \quad \forall i \in N$$

or

$$E_\theta\left[-\left.\frac{\partial C_i}{\partial I_i}\right|_{q_i^*(I_i,\theta)}\right] = 1-\delta \;\; \forall i \in N \tag{17}$$

which are the same as first-best FOCs (see equation (11)).

Thus, we can write:

Proposition 3 *Countries implement the first best, in terms of abatement and in terms of investment, through a series of short-term interactions (pledge-and-implement) regulated by a non-comprehensive but contingent long-term provision.*

That is, despite the weakness of the long-term agreement, which can be abandoned at any time, and even if countries cannot sign an agreement on investments, the first best is attained if the long-term provision defines abatement efforts and transfers that are contingent on the state of nature. The reason for this is that the second and the third terms in (14) disappear because now the share obtained by each country of the additional surplus created through an agreement at stage 4 is no longer under discussion at that stage, since transfers were already determined at stage 1, when the long-term agreement was signed.

As discussed in the introduction, the Paris Agreement is incomplete in a double sense. First, it is non-comprehensive, as it does not determine investment levels. What Proposition 3 tells us is that this incompleteness is not problematic, as the world could attain the first-best by writing a long-term provision that is contingent on the state of nature. In theory, a detailed development of the rulebook, and articles 6 and 9 of the Paris Agreement, could make the long-term provision contingent on the state of nature. However, it is no secret that the practical challenges of this task would be enormous. For this reason I now move on to analyze a long-term provision that is also incomplete in this second sense, i.e. non-contingent.

3.4 Non-comprehensive and non-contingent long-term provision

Assume now that θ is so complex and multidimensional that it is not possible to write a meaningful long-term provision as a function of the state of nature. This is clearly the case for climate change, as the number of variables is extremely large and the uncertainties surrounding them are also large. Thus, countries cannot write down future abatement efforts and associated transfers as a function of the state of nature finally realized.

Let us therefore assume that countries write a long-term provision where abatement efforts and associated transfers are specified independently from the state of nature. This provision states, e.g., that country j has to pay country i a given amount of money for increasing its abatement efforts to a certain level. The difference with the previous section is that the amount paid is determined *a priori*, and that it does not depend on the state of nature finally realized.

Under this assumption, let us fix an arbitrary incomplete long-term provision $\tau = (\tau_1, ..., \tau_N)$. At stage 4, investment levels $\mathbf{I}$ are sunk and the state of nature θ is known. Once θ is observed, all countries are happy with the terms of this agreement if

$$v_i(\mathbf{q}^l, I_i, \theta) + \tau_i \geq v_i(\mathbf{q}^A, I_i, \theta) \;\; \forall i \in N, \tag{18}$$

where l in superscript stands for long-term non-comprehensive and non-contingent agreement. If these conditions hold for all countries, I assume that they honour the agreement, as in the previous sub-section. If (18) does not hold for at least one country (an eventuality that was not possible in the previous sub-section), the long-term agreement is not honoured and countries renegotiate each pledge-and-implement short-term interaction following the two-phases process described in section 2. As before, I assume that countries cannot write a new long-term provision (this assumption again has no strategic bite as the state of nature is known and investments are sunk). That is, in each pledge-and-implement short-term interaction countries will renegotiate and agree on a share that implies a set of transfers τ' such that

$$v_i(\mathbf{q}^l, I_i, \theta) + \boldsymbol{\tau}'_i \geq v_i(\mathbf{q}^A, I_i, \theta) \text{ for } i = 1, ..N.$$

Note that this new sequence of short-term interactions has an expected payoff identical to (13). Hence, at stage 4 equilibrium expected payoffs are defined by the terms given by the long-term provision if they are profitable for all parts, and by the expected payoffs associated with the sequence of short-term interactions without any long-term provision otherwise. That is, the payoff is given by:

$$\begin{cases} \frac{v_i(\mathbf{q}^l, I_i, \theta)+\tau_i}{1-\delta} & \text{if } \tau_i \geq v_i(\mathbf{q}^A, I_i, \theta) - v_i(\mathbf{q}^l, I_i, \theta) \text{ for } i = 1, ..N \\ \frac{v_i(\mathbf{q}^A, I_i, \theta)+\varphi_i \sum_{i\in N}\left[v_i(\mathbf{q}^S, I_i, \theta)-v_i(\mathbf{q}^A, I_i, \theta)\right]}{1-\delta} & \text{otherwise.} \end{cases}$$

Define now, for each vector of investment $\mathbf{I}$, the set

$$\Psi(\mathbf{I}) = \left\{\theta \in \Theta : \ \tau_i \geq v_i(\mathbf{q}^A, I_i, \theta) - v_i(\mathbf{q}^l, I_i, \theta)\right\}.$$

Set $\Psi(\mathbf{I})$ includes all the states of nature, given the investment vector decided at stage 2, for which (18) holds. Hence, given the arbitrary long-term provision $\boldsymbol{\tau}$, the expected payoff at stage 2 for country i is:

$$E_{\theta\in\Psi(\mathbf{I})}\left[\frac{v_i(\mathbf{q}^l, I_i, \theta) + \tau_i}{1-\delta}\right] + E_{\theta\in\Theta/\Psi(\mathbf{I})}\left[\frac{v_i(\mathbf{q}^A, I_i, \theta) + \varphi_i \sum_{i\in N}\left[v_i(\mathbf{q}^S, I_i, \theta) - v_i(\mathbf{q}^A, I_i, \theta)\right]}{1-\delta}\right] - I_i \tag{19}$$

For abatement the FOC is again given by (10), implying $q_i^l(I_i, \theta) = q_i^S(I_i, \theta) = q_i^*(I_i, \theta)$. For investment, applying again the envelope theorem, but taking into account that $\frac{\partial v_j(\mathbf{q}^A, I)}{\partial q_i^A} \neq 0$, the NE investment vector $\mathbf{I}^l$ is given by:

$$\begin{aligned} & E_{\theta\in\Psi(I)}\left[-\left.\frac{\partial C_i}{\partial I_i}\right|_{q_i^*(\cdot)}\right] + \\ & E_{\theta\in\Theta/\Psi(I)}\left[-\varphi_i \left.\frac{\partial C_i}{\partial I_i}\right|_{q_i^*(\cdot)} - (1-\varphi_i)\left.\frac{\partial C_i}{\partial I_i}\right|_{q_i^A(\cdot)} - \sum_{j\in N/i} \varphi_i \left.\frac{\partial B_j}{\partial q_i}\right|_{q_i^A(\cdot)} \frac{q_i^A(I_i)}{\partial I_i}\right] \\ = & \ 1-\delta \quad \text{for } i = 1, ..N \text{ and } j \neq i \end{aligned} \tag{20}$$

Note that the LHS of (20) is a linear combination of (11) and (14), implying that the non-comprehensive and incomplete long-term provision brings the investment level closer to the first best. Thus, we can write the following proposition:

Proposition 4 *Countries do not implement the first best through a series of short term interactions (pledge-and-implement) regulated by a non-comprehensive and non-contingent long-term provision. Given a level of investment, abatement is optimal, as it is defined by (10), i.e.* $q_i^l(I_i, \theta) = q_i^S(I_i, \theta) = q_i^*(I_i, \theta)$. *However, there is underinvestment, although it is less severe than in the absence of the long-term provision. As a consequence, abatement is also below the optimal level but closer to the first best than it would be without the long-term* *provision. Moreover, each additional state of nature considered in the long-term provision, i.e. included in $\Psi(I)$, weakly approaches the solution to the first best. If $\Psi(I) \rightarrow \Theta$, the equilibrium tends to the first best.*

As discussed above, long-term provisions in the Paris Agreement are, at best, non-comprehensive and non-contingent. However, despite this weakness, Proposition 4 shows that a long-term provision has a role to play, although it is not able to bring investment to the optimal level, it brings investments closer to the first best. Propositions 3 and 4 also tell us that being non-comprehensive is not the main weakness of the long-term provision, as investments would be optimal even if countries cannot agree on them, provided that they were able to write a contingent long-term provision, in the sense that all possible states of nature are considered. Furthermore, the more complete the long-term provision becomes in this sense, the closer it moves us to the first best. Thus, the Paris Agreement should define as clearly as possible the conditions under which transfers will occur under Articles 6 and 9 and these conditions should depend as much as possible on the severity of the climate change finally observed (and on other stochastic variables).

4 Discussion

In this section, I relax some of the key assumptions used above, discuss the likely impact of alternative frameworks, and relate the results obtained to real-life negotiations.

4.1 Extensions

4.1.1 Order is not known *a priori*

I have assumed that the order in which countries talk at the implementation phase is known. This is a reasonable assumption because it captures in a stylized manner the relative importance of different countries in the international arena. Alternatively, one could assume that the order is not known and that it will only be determined randomly at the beginning of the implementation phase. Then, during the pledge phase each country would assume that it has an equal opportunity of talking first. As shown in Appendix A.2, this would introduce an inefficiency and the first best could no longer be attained. The reason for this is that countries would submit an 'average' pledge, taking into account that they might talk in any order. However, at least the country finally talking in last position will obtain a smaller than average share, and will therefore suffer costs-of-revoking.

4.1.2 Not all countries make conditional pledges

In the equilibrium discussed above all countries make conditional pledges. Let us now discuss why this is the case and what would happen if not all countries made conditional pledges. Assuming that a country does not submit a conditional pledge is equivalent to assuming that its cost-of-revoking is zero. From (6) we know that any player i prefers the situation where $k_i > 0$, i.e. it submits a conditional pledge, while for all other players submit no conditional pledge, i.e. $k_j = 0$ for all $j \neq i$, as this would maximize the share obtained by country i of the additional surplus created. As this holds for all players, the situation is akin to a prisoner's dilemma and, hence, all countries end up submitting conditional pledges in equilibrium. When the order in which countries talk at the implementation phase is known (as in the main analysis), the fact that all countries submit conditional pledges has no consequences in terms of overall welfare, as the short-term first best is anyway attained. However, when the order in which countries talk at the implementation phase is not known *a priori*, as discussed in the previous sub-section, this yields a situation that is Pareto-dominated by the alternative where no side makes conditional pledges.

4.2 Key assumptions and further research

4.2.1 Stationarity

The analysis above focuses on SSPE. This is a common assumption in applied multilateral bargaining games. The reason is that, as shown by Shaked and reported by Sutton (1986), the generalization of Rubinstein's bilateral bargaining process to a multilateral framework yields multiple equilibria when unanimity is required and strategies are not stationary. There are different alternatives to restore uniqueness even with non stationary strategies. A straightforward approach is to assume a terminal period T for the negotiations, as the model can then be solved by backward induction. However, even within a infinite horizon bargaining model there are alternatives, such as the possibility of making conditional and unconditional offers financed by outside money (Huang, 2002), which has been extended to games with externalities in Caparrós and Péreau (2017). Huang (2002) shows that, as long as all players are allowed to talk once following a cyclical protocol, the solution of the game allowing for non-stationary strategies converges to the SSPE of the unanimity game described above when $\delta \to 1$ (the same result is obtained with the 'exit-game' proposed by Krishna and Serrano 1996). Although I have left a formal analysis for further research, the results shown above would almost certainly continue to hold under these alternative assumptions.

4.2.2 Stock variable for climate change.

The analysis above has not considered the impact of including a stock variable to capture CO_2 accumulation in the atmosphere. Battaglini and Harstad (2016) or Karp and Sakamoto (2019) are recent examples that do incorporate climate stocks in their dynamic analyses. However, the key assumption in these analyses is the use of the linearity-in-state to re-write the stock dependent problem as a related problem where the utility is independent of all past stocks (see, e.g., the Lemma 1 in Battaglini and Harstad 2016). As detailed in Karp and Sakamoto (2019), see their Propositions 4.1 and 4.2, by keeping the assumption of linearity in state it is possible to build a structural climate change game with a stock state variable and a infinite time horizon in which any equilibrium of the corresponding reduced-form model is also an equilibrium of the structural model, although there may be equilibria of the structural model with $T = \infty$ that are not equilibria of the reduced form model. Thus, although I have decided to leave a formal analysis out of this paper, the equilibrium of the model discussed above (seen as a reduced-form model) would likely be an equilibrium of a properly defined structural isomorphic model, as long as the linearity in state assumption is maintained.

4.3 Relating the results to real-life negotiations

My interpretation of the current situation of the Paris Agreement is as follows. Most countries of the world (approximately 80%, see the introduction) have submitted conditional pledges, in addition to their unconditional pledges. As shown above, this is an appropriate strategy as it can potentially bring the world closer to the short-term first-best, for a given level of investment. The remaining 20% of countries are developed countries that have submitted, strictly speaking, only unconditional pledges. However, as they will soon (in principle by 2025, see the introduction) submit pledges for their contribution to the transfers to developing countries, one could consider those as conditional pledges.[10] Thus, soon all countries which are part of the Paris Agreement[11] will have submitted some form of conditional pledges. This is consistent with the results discussed in section 4.1.2 that all countries have an interest in submitting conditional pledges.

[10] In the Cancun Agreements (a precedent to the Paris Agreement) the EU submitted another form of conditional pledge, stating a more ambitious "conditional offer [...] provided that other developed countries commit themselves to comparable emission reductions", see FCCC/SBSTA/2014/INF.6, available at https://unfccc.int. My analysis shows that it was probably a bad idea to abandon this path in the Paris Agreement, as developed countries should focus more on conditional pledges, in their own interest.

[11] As discussed in Mehling *et al.* (2018), the US could potentially participate in my 'implementation phase', what they call 'mitigation output transfers', even if they are an outsider to the Paris Agreement. My analysis suggests that the US should announce a form of conditional partial commitment (pledge), even if it does so outside of the Paris Agreement.

This can also serve as an explanation for the fact that almost all the countries of the world have submitted pledges, while only a small part of them (the developed countries) accepted commitments as part of the Kyoto Protocol. Once some countries submit conditional pledges, it is in the interest of all the others to follow suit. Then, once all countries have submitted their pledges, they are all interested in entering the negotiations at the implementation phase (to obtain a share of the overall additional surplus, but also to avoid individual costs-of-revoking). Hence, from a policy perspective, compared to the direct approach in the Kyoto Protocol, the Paris Agreement has the advantage of attracting more negotiators to the table.

I have also shown that the fact that all countries submit conditional pledges could lead to a Pareto-dominated situation compared to the case without pledges. However, this is only true if the order in which countries talk at the implementation phases is not known *a priori* (section 4.1.1). As this order is meant to capture in a stylized manner the relative importance of the different countries at the international arena, and this is relatively well known and stable, the assumption of a known order is more adequate.

In any case, even if the pledge-and-implement process were fully developed, it cannot prevent the long-term underinvestment problem. The long-term provisions in the current version of the Paris Agreement are probably not sufficiently developed to qualify as a long-term provision in the sense discussed above. Although there is a long-term goal, a floor for the intended transfers from developed countries (USD 100 billion a year), and a request to countries to submit their long-term mid-century strategies (see the introduction), at the current state of affairs countries have not determined sufficiently future pledges and associated transfers, either without (non-contingent) or with (contingent) a reference to the underlying uncertainty. This is problematic, as it implies that long-term investments will not be optimal according to my analysis. Recent developments, such as the Paris rule-book and the Katowice package are steps in the right direction, but the conditions for future transfers should be specified, to deter incentives to hold-up investments. The analysis has also shown that being non-comprehensive is not the main weakness of a potential long-term provision, as investments would be optimal even if countries cannot agree on them, as long as they were able to write a contingent agreement, in the sense that all possible states of nature are considered. Furthermore, even 'real world' long-term provisions, which by definition will not be fully contingent, can bring the world closer to the first-best and should therefore be pursued. As discussed above, this implies that the Paris Agreement should condition future transfers under Articles 6 and 9 on relevant stochastic variables, such as observed climate change and economic conditions.

5 Conclusion

This paper presents a stylized model to analyze the two two-tier process initiated by the Paris Agreement, where a long-term non-binding agreement is followed by a series of short-term interactions governed by a pledge-and-implement mechanism. The pledge-and-implement mechanism has been modeled as a bargaining game in which countries submit conditional and unconditional pledges, which are not binding and not subject to approval by other countries. Then, eventually, countries bargain over the conditions set out in the conditional pledges. However, backing-off from the initial terms proposed in the conditional pledges has reputational costs.

The paper also analyzes long-term climate provisions that are incomplete in two senses. Firstly, they are non-comprehensive because countries do not agree on investments, only on rules governing future short-term interactions on abatement levels and associated transfers. Secondly, long-term provisions on abatement are incomplete because the climate change problem is so complex and multidimensional that it is not possible to write a meaningful agreement as a function of the state of nature. Hence, transfers are non-contingent and cannot depend on the state of nature finally observed.

The results show that the pledge and implement mechanism can implement the first-best in the short-term, and that the surplus is shared according to the relative importance of the cost-of-revoking the pledges. However, without a long-term provision there is underinvestment. Even if the agreement does not cover investments, a long term provision contingent on the state of nature can bring the world to the long-term first-best, and a non-contingent provision can bring it closer. The bottom line is that, if fully developed, the Paris Agreement could turn out to be a good instrument. However, the analysis also shows that the current Paris Agreement needs to be further developed.

References

[1] Barrett, S. 1994. Self-enforcing international environmental agreements. *Oxford Econ Pap* 46: 878-894.

[2] Barrett, S., 2002. Consensus Treaties. *J Inst Theor Econ* 158: 529-547.

[3] Barrett, S. and Dannenberg, A., 2016. An experimental investigation into ‘pledge and review’ in climate negotiations. *Clim change* 138: 339-351.

[4] Battaglini, M. and Harstad, B., 2016. Participation and Duration of Environmental Agreements. *J Political Econ* 124(1): 160-204.

[5] Beccherle, J. and Tirole, J., 2011. Regional Initiatives and the Cost of Delaying Binding Climate Change Agreements. *J Public Econ* 95 (11-12): 1339-1348.

[6] Caparrós, A., 2016. Bargaining and International Environmental Agreements. *Environ Resour Econ* 65(1): 5–31.

[7] Caparrós, A. and Péreau, J.C., 2017. Multilateral versus sequential negotiations over climate change. *Oxford Econ Pap* 69(2): 365–387.

[8] Caparrós, A., Péreau, J.-C. and Tazdaït, T., 2004. North-South Climate Change Negotiations: a Sequential Game with Asymmetric Information. *Public Choice* 121 (3-4): 455-480.

[9] Crawford, V., 1988. Long-Term Relationships Governed by Short-Term Contracts. *Am Econ Rev* 78: 485-499

[10] Dutta, P.K. and Radner, R. 2004. Self-enforcing Climate-Change Treaties. *Proc Natl Acad Sci USA* 101: 4746–4751.

[11] Dutta, P.K. and Radner, R., 2009. A Strategic Analysis of Global Warming: Theory and Some Numbers. *J Econ Behavr Organ* 71: 187–209.

[12] Dutta, P.K. and Radner, R. 2012. Capital Growth in a Global Warming Model: Will China and India Sign a Climate Treaty? *Econ Theory* 49: 411–443.

[13] Dutta, P.K. and Radner, R. 2019. *The Paris Accord Can Be Effective if the Green Climate Fund is Effective*. Mimeo.

[14] Finus, M. and Caparrós, A., 2015. *Game Theory and International Environmental Cooperation: Essential Readings*. Edward Elgar, Cheltenham.

[15] Harstad, B., 2012. Climate Contracts: A Game of Emissions, Investments, Negotiations, and Renegotiations. *Rev Econ Stud* 79(4): 1527-57.

[16] Harstad, B., 2016. The Dynamics of Climate Agreements. *J Euro Econ Assoc* 14(3): 719–752.

[17] Harstad, B., 2019. *Pledge-and-Review Bargaining: From Kyoto to Paris*. Mimeo.

[18] Harstad, B., 2020. *A theory of pledge-and-review bargaining*. Mimeo.

[19] Hart, O. and Holstrom, B., 1987. The Theory of Contract. In T. Bewley (ed.) *Advances in Economic Theory*. Cambridge University Press.

[20] Hart, O. and Moore, J., 1988. Incomplete Contracts and Renegotiation. *Econometrica* 56: 755-786.

[21] Huang, C.-Y., 2002. Multilateral bargaining: conditional and unconditional offers. *Econ Theory* 20: 401–412.

[22] Jotzo, F., Depledge, J. and Winkler, H. 2018. US and international climate policy under President Trump. *Clim Policy* 18(7): 813-817.

[23] Karp, L. and Sakamoto, H., 2019. Sober optimism and the formation of international environmental agreements. *Kyoto University Discussion Paper* No. E-19-002.

[24] Krishna, V. and Serrano, R., 1996. Multilateral bargaining. *Rev Econ Stud* 63:61–80.

[25] Leventoglu, B. and Tarar, A., 2005. Prenegotiation Public Commitment in Domestic and International Bargaining. *Am Polit Sci Rev* 99(3): 419-433.

[26] Masson-Delmotte, V., *et al.* (eds.), 2018.*Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strngthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty.* Available at https://www.ipcc.ch/sr15

[27] Mehling, M.A., Metcalf, G.E. and Stavins, R.N., 2018. Linking climate policies to advance global mitigation. *Science* 359(6379): 997-998.

[28] Muthoo, A., 1995. Bargaining in a Long-Term Relationship with Endogenous Termination. *J Econ Theory* 66: 590-598.

[29] Muthoo, A., 1996. A Bargaining Model Based on the Commitment Tactic. *J Econ Theory* 69: 134-152.

[30] Nordhaus, W., 2019. Climate Change: The Ultimate Challenge for Economics. *Am Econ Rev* 109(6): 1991-2014.

[31] Rogelj, J., Elzen, M., Höhne, N., Fransen, T., Fekete, H., Winkler, H., Schaeffer, R., Sha, F., Riahi, K. and Meinshausen, M., 2016. Paris Agreement climate proposals need a boost to keep warming well below 2°C. *Nature* 534: 631-639.

[32] Segal, I. and Whinston, M. D., 2010. Property Rights. In Gibbons, R. and Roberts, J. (eds) *Handbook of Organizational Economics*. Princeton University Press.

[33] Schneider, L., Duan, M., Stavins, R., Kizzier, K., Broekhoff, D., Jotzo, F., Winkler, H., Lazarus,M., Howard, A. and Hood, C., 2019. Double counting and the Paris Agreement rulebook. *Science* 80(366): 180–183.

[34] UNFCCC, 2015. *Adoption of the Paris Agreement* (FCCC/CP/2015/L.9/ Rev.1). Available at unfcc.int.

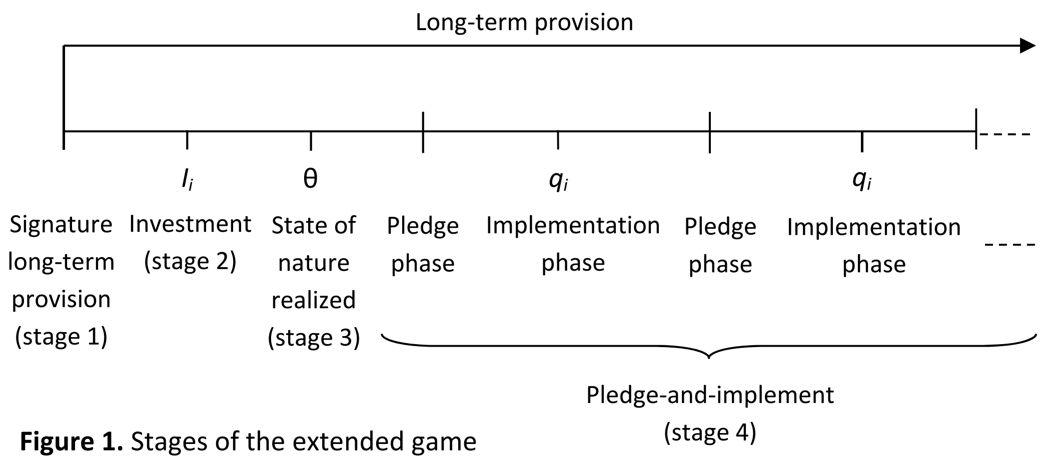

Figure 1. Stages of the extended game

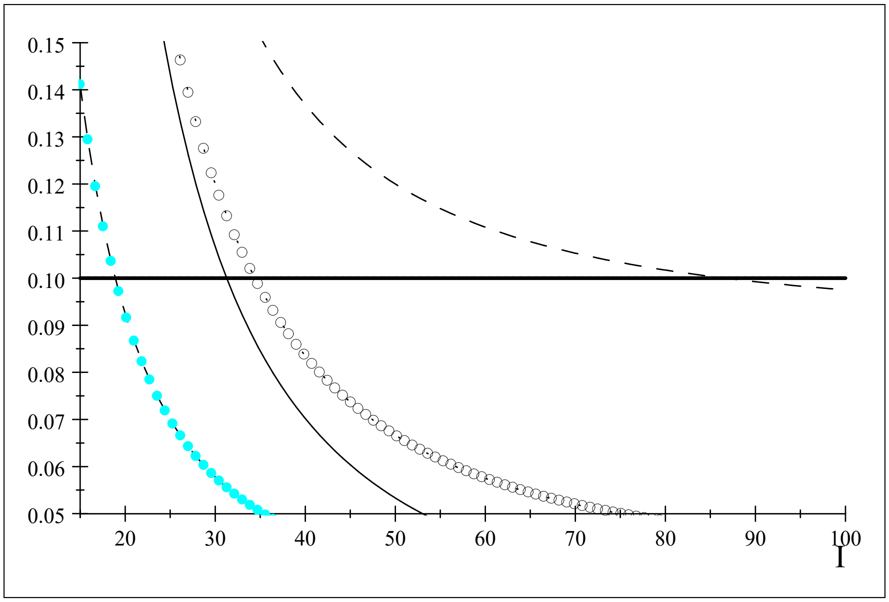

Figure 2. Investment in the first-best and in the 'basic Paris Agreement', using functional forms and parameter values of Example 1. Thick solid: RHS of (11) and (14), i.e. $1 - \delta$. Dashed: LHS of (11). Dots: first term LHS of (14). Circles: first two terms LHS of (14). Thin solid: LHS of (14). Intersection of dashed and thick solid: I^*. Intersection of thin and thick solid: I^S.

A Proofs

A.1 Proposition 1

From Lemma 1 we know that unconditional pledges are defined by the abatement level of the Nash equilibrium without interactions, i.e. $\mathbf{q}^U = \mathbf{q}^A$. We also know that equilibrium abatement efforts in the eventuality of an agreement based on the conditional pledges are first-best, i.e. $\mathbf{q}^C = \mathbf{q}^*$, as countries have no interest in leaving unexhausted surplus in the case of an agreement, i.e. the additional surplus created is $\Delta = \Delta(\mathbf{q}^A, \mathbf{q}^*)$. I focus in the remainder of the proof on the case with three players. Nevertheless, the same strategy can be applied to any number of countries. Thus, I label countries as 1, 2 and 3.

In any equilibrium we must have that $\sum_{j=1}^{3} \varphi_i = 1$. If the shares add up to more than one the equilibrium cannot be implemented and if they add up to less than one part of the surplus would not be appropriated by any country. Nevertheless, note that there is no obvious reason why $\sum_{j=1}^{3} \alpha_i$ should add up to one, although I will show below that this is the case in equilibrium. At the implementation phase, phase 2, the conditional NDCs are already know, so that α_1, α_2 and α_3 are given. Equilibrium offers at the implementation phase have to meet the following relationships. At time 3, country 1 offers a share φ_1^3 for itself, a share φ_2^3 for country 2 and a share $(1 - \varphi_1^3 - \varphi_2^3)$ for country 3 (superscripts indicate the period when the offer is made). This provides a payoff equal to $\varphi_1^3 - k_1(\alpha_1 - \varphi_1^3)$ for country 1, equal to $\varphi_2^3 - k_2(\alpha_2 - \varphi_2^3)$ for country 2 and equal to $(1 - \varphi_1^3 - \varphi_2^3) - k_3(\alpha_3 - (1 - \varphi_1^3 - \varphi_2^3))$ for country 3. In the previous period (time 2), country 3 makes an offer that makes countries 1 and 2 indifferent between accepting it or waiting one period for the offer of country 1. Thus, the share offered to country 1, denoted φ_1^2, has to ensure a payoff for 1 equal to $\delta\left(\varphi_1^3 - k_1(\alpha_1 - \varphi_1^3)\right)$. The share offered to country 2 by country 3, φ_2^2, has to provide a payoff equal to $\delta\left(\varphi_2^3 - k_2(\alpha_2 - \varphi_2^3)\right)$. This leaves country 3 with a share $1 - \varphi_1^2 - \varphi_2^2$ and a payoff $(1 - \varphi_1^2 - \varphi_2^2) - k_3(\alpha_3 - (1 - \varphi_1^2 - \varphi_2^2))$. At time 1, country 2 makes an offer that leaves 1 and 3 indifferent between accepting or waiting for the next period equilibrium offer. That is, country 2 offers them a share φ_1^1 and φ_3^1, respectively, yielding payoffs equal to $\delta^2\left(\varphi_1^3 - k_1(\alpha_1 - \varphi_1^3)\right)$ and $\delta\left((1 - \varphi_1^2 - \varphi_2^2) - k_3(\alpha_3 - (1 - \varphi_1^2 - \varphi_2^2))\right)$ for countries 1 and 3. This leaves a share equal to $1 - \varphi_1^1 - \varphi_3^1$ and a payoff equal to $(1 - \varphi_1^1 - \varphi_3^1) - k_2(\alpha_2 - (1 - \varphi_1^1 - \varphi_3^1))$ for country 2. Finally, at time 0 country 1 makes an offer that leaves 2 and 3 indifferent between accepting or rejecting. That is, country 1 offers shares φ_2^0 and φ_3^0, yielding a payoff of $\delta\left((1 - \varphi_1^1 - \varphi_3^1) - k_2(\alpha_2 - (1 - \varphi_1^1 - \varphi_3^1))\right)$ for country 2 and a payoff of $\delta^2\left((1 - \varphi_1^2 - \varphi_2^2) - k_3(\alpha_3 - (1 - \varphi_1^2 - \varphi_2^2))\right)$ for country 3. This leaves country 1 with a share equal to $1 - \varphi_2^0 - \varphi_3^0$ and a payoff equal to $(1 - \varphi_2^0 - \varphi_3^0) - k_1(\alpha_1 - (1 - \varphi_2^0 - \varphi_3^0))$.

Assuming stationarity, we know that the offers made by country 1 at time 0 and time 3 have to be identical, i.e. $\varphi_1^0 = \varphi_1^3 = \varphi_1^*$, and yield identical payoffs for countries 2 and 3, i.e.:

$$\begin{aligned} \varphi_2^3 - k_2(\alpha_2 - \varphi_2^3) &= \delta\left(\left(1-\varphi_1^1-\varphi_3^1\right)(1+k_2) - k_2\alpha_2\right) \\ \left(1-\varphi_1^3-\varphi_2^3\right)(1+k_3) - k_3\alpha_3 &= \delta^2\left(\left(1-\varphi_1^2-\varphi_2^2\right)(1+k_3) - k_3\alpha_3\right) \end{aligned} \tag{A1}$$

From the discussion above we also know that the offers φ_1^2, φ_2^2, φ_1^1 and φ_3^1 have to ensure that the following equalities hold:

$$\begin{aligned} \delta(\varphi_1^3 - k_1(\alpha_1 - \varphi_1^3)) &= \varphi_1^2 - k_1(\alpha_1 - \varphi_1^2) \\ \delta\left(\varphi_2^3 - k_2(\alpha_2 - \varphi_2^3)\right) &= \varphi_2^2 - k_2(\alpha_2 - \varphi_2^2) \\ \delta^2\left(\varphi_1^3 - k_1(\alpha_1 - \varphi_1^3)\right) &= \varphi_1^1 - k_1(\alpha_1 - \varphi_1^1) \\ \delta\left(\left(1-\varphi_1^2-\varphi_2^2\right) - k_3(\alpha_3 - \left(1-\varphi_1^2-\varphi_2^2\right))\right) &= \varphi_3^1 - k_3(\alpha_3 - \varphi_3^1) \end{aligned} \tag{A2}$$

Substituting the equations in (A2) into the equations in (A1) yields

$$\varphi_1 = \frac{\begin{array}{c} 1 + \left(1+\alpha_1\left(\delta+\delta^2\right)\right)k_1 + (1-\alpha_2)\,k_2 + (1-\alpha_3)\,k_3 \\ + \left(1-\alpha_2+\alpha_1\left(\delta+\delta^2\right)\right)k_1k_2 + \left(1+\alpha_1\left(\delta+\delta^2\right)-\alpha_3\right)k_1k_3 \\ + (1-\alpha_2-\alpha_3)\,k_2k_3 + \left(1-\alpha_2-\alpha_3+\alpha_1\left(\delta+\delta^2\right)\right)k_1k_2k_3 \end{array}}{(k_1+1)(k_2+1)(k_3+1)\left(\delta^2+\delta+1\right)} \tag{A3}$$

$$\varphi_2 = \frac{\begin{array}{c} \delta + (\delta-\alpha_1\delta)\,k_1 + \left(\alpha_2+\delta+\alpha_2\delta^2\right)k_2 + (\delta-\alpha_3\delta)\,k_3 \\ + \left(\delta-\alpha_1\delta+\alpha_2\left(1+\delta^2\right)\right)k_1k_2 + (\delta-\alpha_1\delta-\alpha_3\delta)\,k_1k_3 \\ + \left(\delta-\alpha_3\delta+\alpha_2\left(1+\delta^2\right)\right)k_2k_3 + \left(\delta-\alpha_1\delta-\alpha_3\delta+\alpha_2\left(1+\delta^2\right)\right)k_1k_2k_3 \end{array}}{(k_1+1)(k_2+1)(k_3+1)\left(\delta^2+\delta+1\right)} \tag{A4}$$

To be an equilibrium, for these values it must be true that $\varphi_1 \leq \alpha_1$, $\varphi_2 \leq \alpha_2$ and $1-\varphi_1-\varphi_2 \leq \alpha_3$. First, I show that there is an equilibrium where these inequalities become equalities. Then I show that this is the unique equilibrium. The system formed by (A3)-(A4) and $\varphi_1 = \alpha_1$, $\varphi_2 = \alpha_2$ and $1-\varphi_1-\varphi_2 = \alpha_3$ has the following solution:

$$\begin{aligned} \alpha_1 &= \varphi_1^* = \frac{(1+k_1)}{(1+k_1)+\delta(1+k_2)+\delta^2(1+k_3)} = \frac{(1+k_1)}{\sum_{i=1}^3 \delta^{i-1}(1+k_i)} \\ \alpha_2 &= \varphi_2^* = \frac{\delta(1+k_2)}{(1+k_1)+\delta(1+k_2)+\delta^2(1+k_3)} = \frac{\delta(1+k_2)}{\sum_{i=1}^3 \delta^{i-1}(1+k_i)} \\ \alpha_3 &= 1-\varphi_1^*-\varphi_2^* = \frac{\delta^2(1+k_3)}{(1+k_1)+\delta(1+k_2)+\delta^2(1+k_3)} = \frac{\delta^2(1+k_3)}{\sum_{i=1}^3 \delta^{i-1}(1+k_i)} \end{aligned} \tag{A5}$$

To see that a larger α_1 cannot be chosen by country 1 substitute (A3) and (A4) into the payoff function (5), see the main text, to obtain:

$$\Pi_1(\alpha_1, \varphi_1) = \Omega - \frac{k_1\Delta}{\delta^2+\delta+1}\alpha_1,$$

where Ω collects the terms that depend on $\delta, \alpha_2, \alpha_3, d, v_1(q^A), k_1, k_2$ and k_3 but are independent from α_1.

As can be seen, the payoff is a decreasing function of α_1 and country 1 is therefore not interested in proposing a lager α_1. I now show, *a contrario*, that country 1 is also not interested in proposing a smaller α_1. To see this, assume that country 1 proposes $(\alpha_1 - \varepsilon)$. Rewriting (A2) and (A1) for $(\alpha_1 - \varepsilon)$, substituting as above and solving yields a share for country 1 equal to $\hat{\varphi}_1$, using the hat to indicate values for this deviation, such that

$$\begin{aligned}\hat{\varphi}_1 &= \varphi_1 - \frac{(1+\delta)\delta k_1}{(1+k_1)\left(\delta^2+\delta+1\right)}\varepsilon = \varphi_1 - \Upsilon\varepsilon \\ \text{with } 0 &\leq \Upsilon = \frac{(1+\delta)\delta k_1}{(1+k_1)\left(\delta^2+\delta+1\right)} \leq 1\end{aligned}$$

Comparing the payoffs for country 1 without and with the deviation, we have

$$\begin{aligned}\Pi_1(\alpha_1, \varphi_1) &= v_1(q^A) + \varphi_1\Delta - \max(0, k_1(\alpha_1 - \varphi_1)\Delta) \\ \hat{\Pi}_1(\alpha_1, \varphi_1) &= v_1(q^A) + (\varphi_1 - \Upsilon\varepsilon)\,\Delta - \max(0, k_1((\alpha_1 - \varepsilon) - (\varphi_1 - \Upsilon\varepsilon))\Delta)\end{aligned}$$

As $\alpha_1 = \varphi_1$ and $(\alpha_1 - \varepsilon) < (\varphi_1 - \Upsilon\varepsilon) = (\alpha_1 - \Upsilon\varepsilon)$ because $0 \leq \Upsilon \leq 1$, this simplifies to

$$\begin{aligned}\Pi_1(\alpha_1, \varphi_1) &= v_1(q^A, \cdot) + \varphi_1\Delta \\ \hat{\Pi}_1(\alpha_1, \varphi_1) &= v_1(q^A, \cdot) + (\varphi_1 - \Upsilon\varepsilon)\,\Delta\end{aligned}$$

and, as $0 \leq \Upsilon$ and $\varepsilon > 0$, we have that $\hat{\Pi}_1(\alpha_1, \varphi_1) - \Pi_1(\alpha_1, \varphi_1) < 0$. Hence, the deviation is not profitable. A similar argument can be developed for the other countries.

The solution shown in the Proposition generalizes the result in (A5) for N countries.

A.2 Order is not known

The analysis of the implementation phase, second phase, is identical as for Proposition 1. Focusing again on the three countries case, by analogy with the analysis in Appendix A.1, we know that the expected share at the implementation phase for country j if the order is not know is:

$$E(\varphi_j) = \frac{1}{3}\left(\varphi_{j1} + \varphi_{j2} + \varphi_{j3}\right) = \frac{(1+k_j)\left(1+\delta+\delta^2\right)}{3\sum_{j=1}^{N}\delta^{i-1}(1+k_i)}$$

where φ_{ji} stands for the share obtained by country j if it finally talks in position i. Following a similar argument as for Proposition 1, to minimize costs-of-revoking country j sets $\alpha_j = E(\varphi_j)$ at the pledge phase. However, if country j finally talks in third position, it has positive costs of revoking because

$$\alpha_j = E(\varphi_j) > \frac{\delta^2 (1 + k_3)}{\sum_{i=1}^{3} \delta^{i-1} (1 + k_i)} = \varphi_j^*,$$

i.e. the share obtained in equilibrium is smaller than the requested share.